

Computation Institute

Grid Technology Transforming Healthcare

Jonathan C. Silverstein, MD, MS, FACS, FACMI

with collaborative assistance by

Stephan Erberich, PhD

Ian Foster, PhD

Carl Kesselman, PhD

Ravi Madduri

Patrick Soon-Shiong, MD

Steve Tuecke

Grid Technology Transforming Healthcare

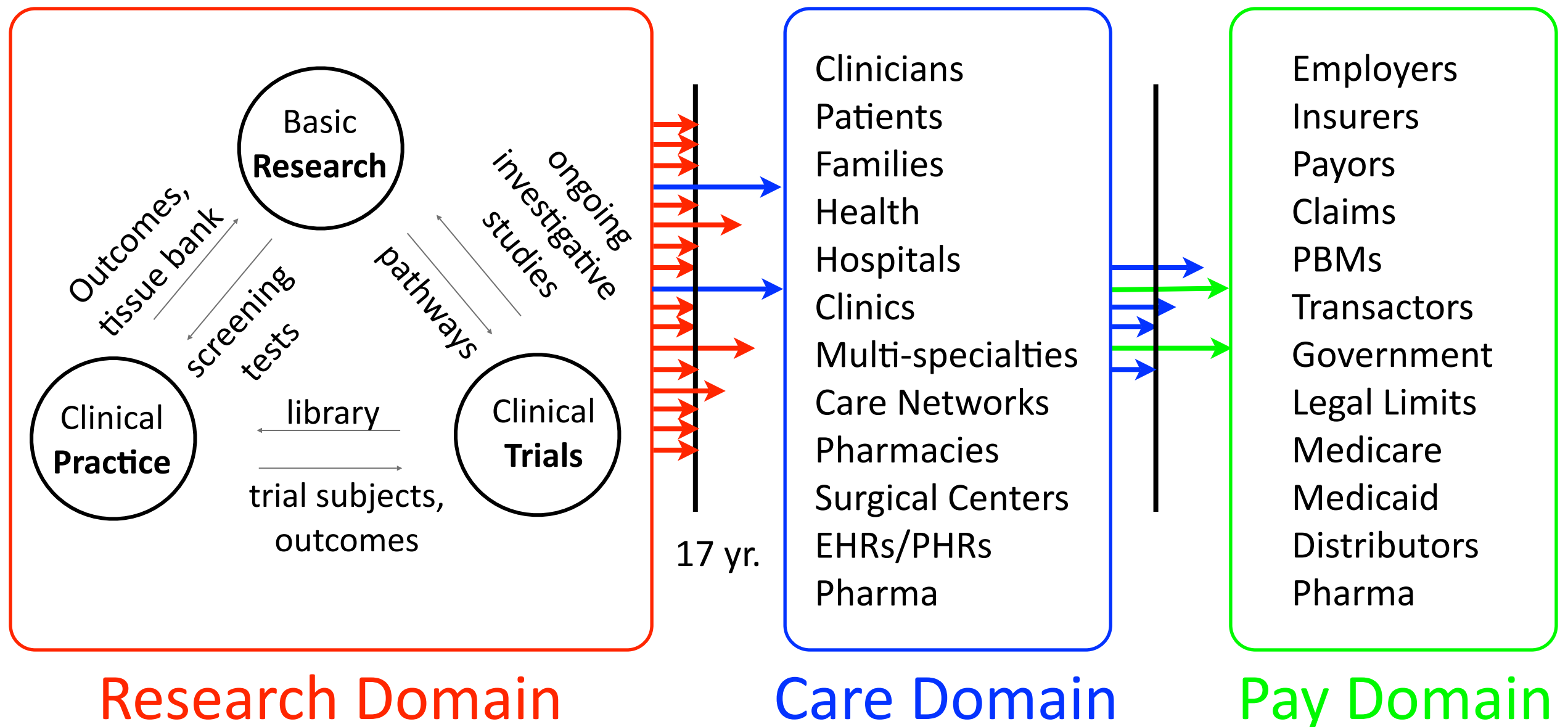
Healthcare in the United States is a complex adaptive system. Individual rewards and aspirations drive behavior as each stakeholder interacts, self-organizes, learns, reacts, and adapts to one another. Having such a system for health is not inherently bad if incentives are appropriately aligned. However, clinical care, public health, education and research practices have evolved into such a state that we are failing to systematically deliver measurable quality at acceptable cost. From a systems level perspective, integration, interoperability, and secured access to biomedical data on a national scale and its creative re-use to drive better understanding and decision-making are promising paths to transformation of healthcare from a broken system into a high performing one. This session will survey HealthGrid issues and projects across clinical care, public health, education and research with particular focus on transformative efforts enabled by high performance computing.

Grid Technology Transforming Healthcare

- Healthcare is a complex ecosystem under tremendous evolutionary pressure
- Federated infrastructure is critical to its survival
- High performance and distributed computing communities are playing a key role but more needs to be done

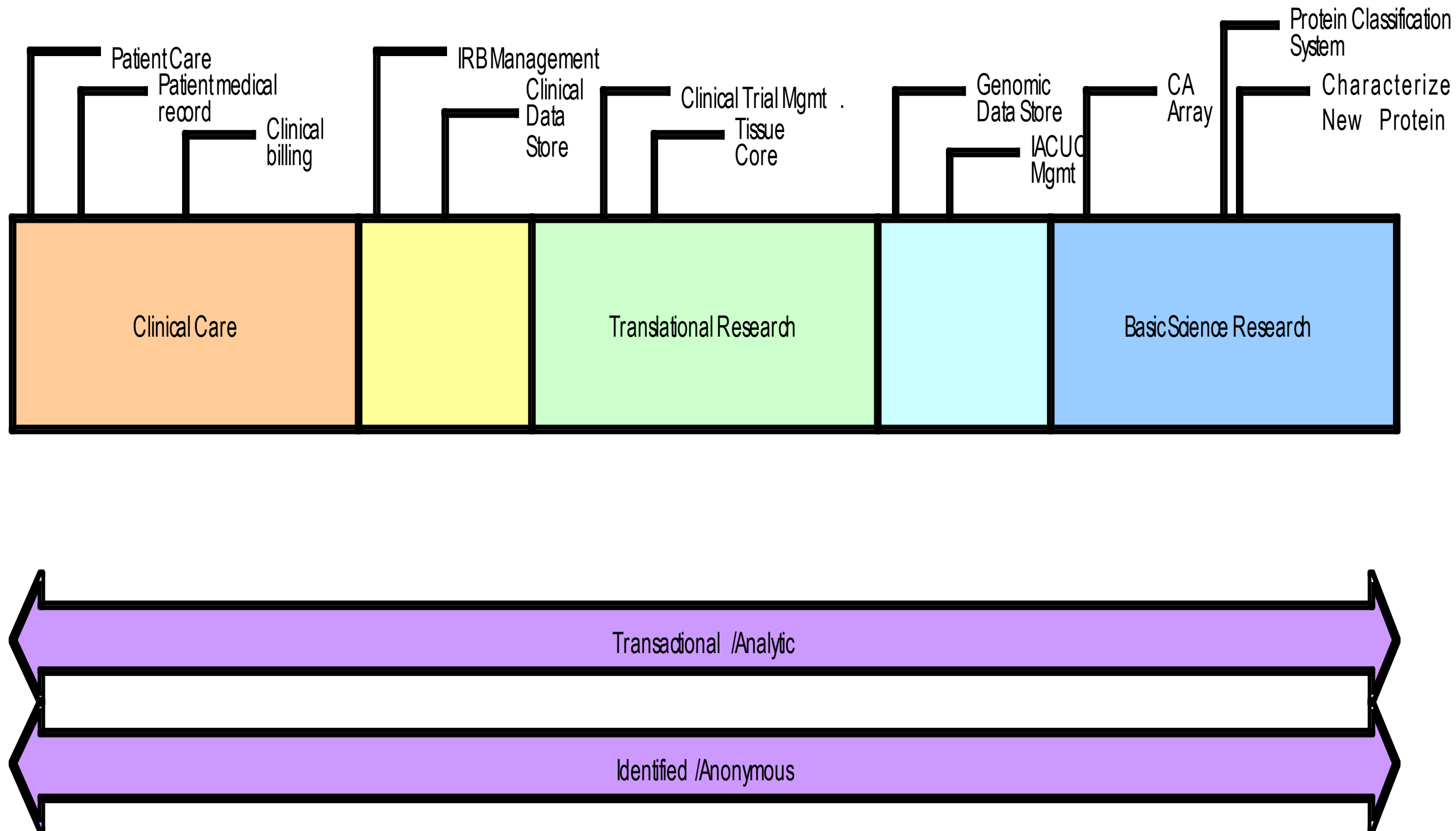
Healthcare Ecosystem Weakly Connected

This extends ideas of healthcare ecosystem by Denis Cortese (Mayo Clinic) and Patrick Soon-Shiong (NCHI)



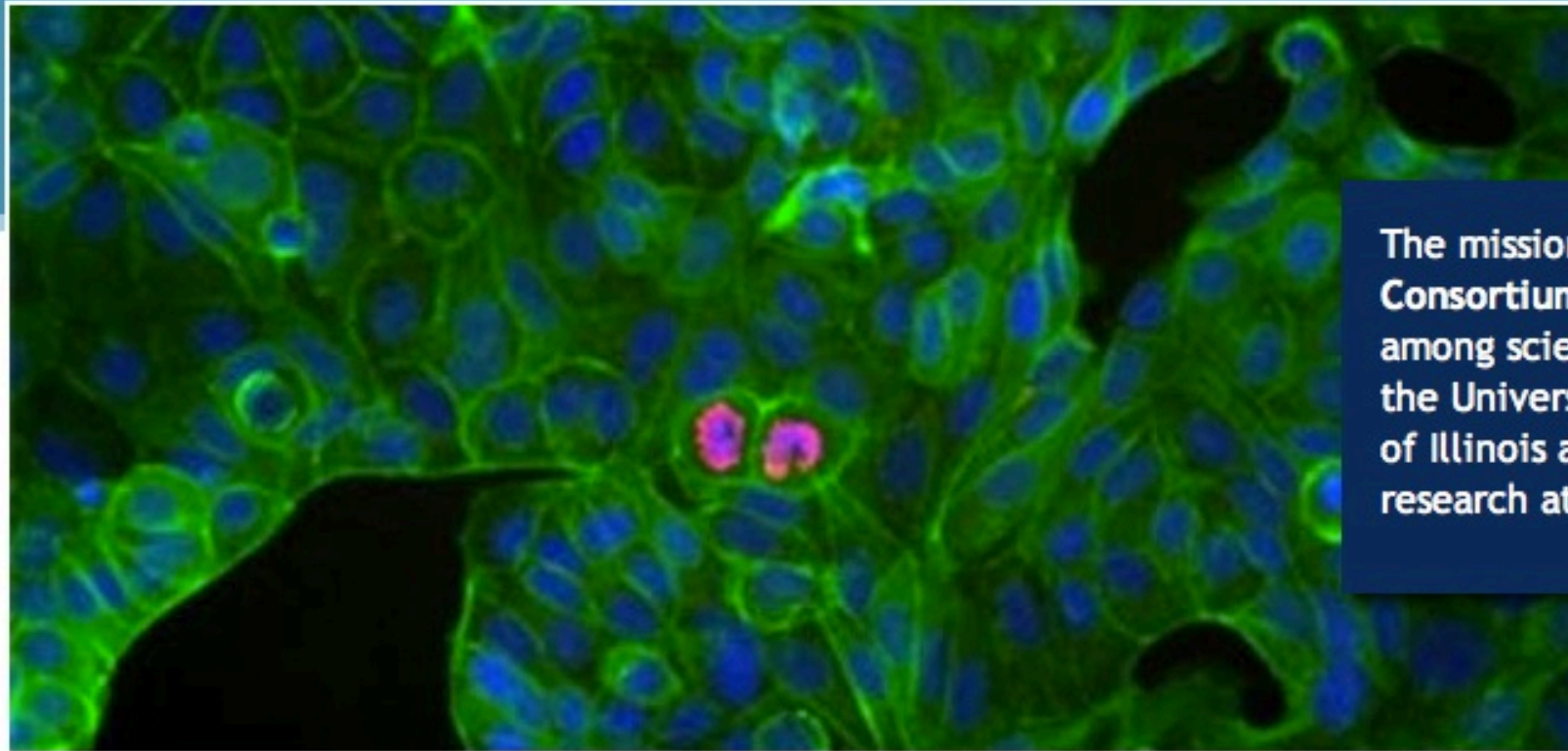
Underpinned by inefficient systems based on ineffective policy incentives

Biomedical Data Stores





CHICAGO BIOMEDICAL CONSORTIUM

[Home](#)[About Us](#)[Grants](#)[Facilities](#)[Education](#)[People](#)[Publications](#)[News](#)[Events](#)[Contact Us](#)

The mission of the **Chicago Biomedical Consortium** is to stimulate collaboration among scientists at Northwestern University, the University of Chicago, and the University of Illinois at Chicago that will transform research at the frontiers of biomedicine.

Popular Links

[RFA: Catalyst](#)[RFA: Spark](#)[CBC Proteomics](#)[Awards](#)[Catalyst Proposal Review](#)

Events

November 19, 2009

UChicago MGCB Seminar series

Genome wide Meiotic Crossover Regulation in *C. elegans*

David Mets

November 19, 2009

NU Neurobiology & Physiology Seminar

Title TBA

DR. Josh Huang, Cold Spring Harbor Labs

November 19, 2009

UIC Cancer Center Research Seminar Series

17 betaHSD-7, a Novel Multifaceted Enzyme that Links the Ovary, Breast Cancer and Brain Development

Geula Gibori, PhD, UIC Distinguished Professor, University of Illinois at Chicago, College of Medicine, Department of Physiology and Biophysics

Highlights

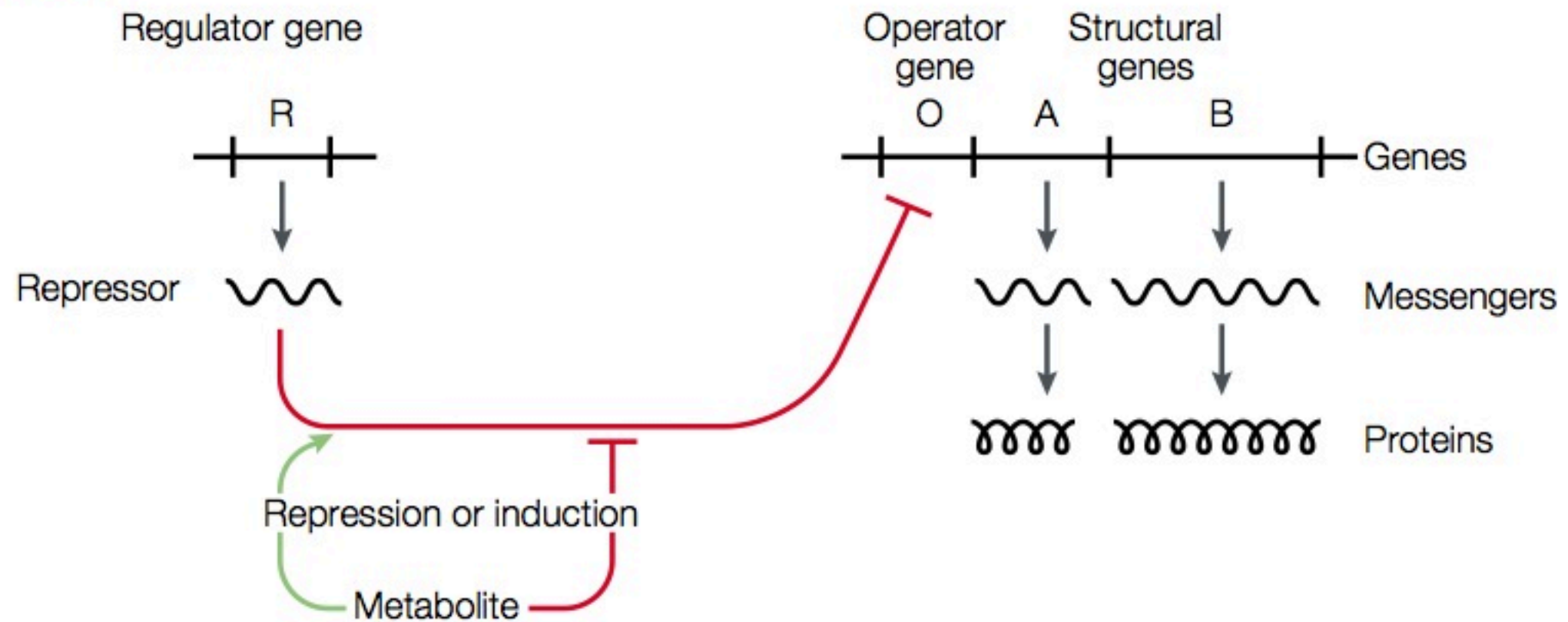
CBC in the News

NATURE | Vol 461 | 22 October 2009

Seeds of collaboration. Academic and government labs in the Chicago area are combining forces to reel in a host of large collaborative research projects — and tens of millions of dollars in funding.



Model I



Jacob and Monod's proposal for the nature of "regulatory genes". Figure 6 in their 1961 paper

NON-CODING RNA GENES AND THE MODERN RNA WORLD

Sean R. Eddy

926 | DECEMBER 2001 | VOLUME 2

www.nature.com/reviews/genetics

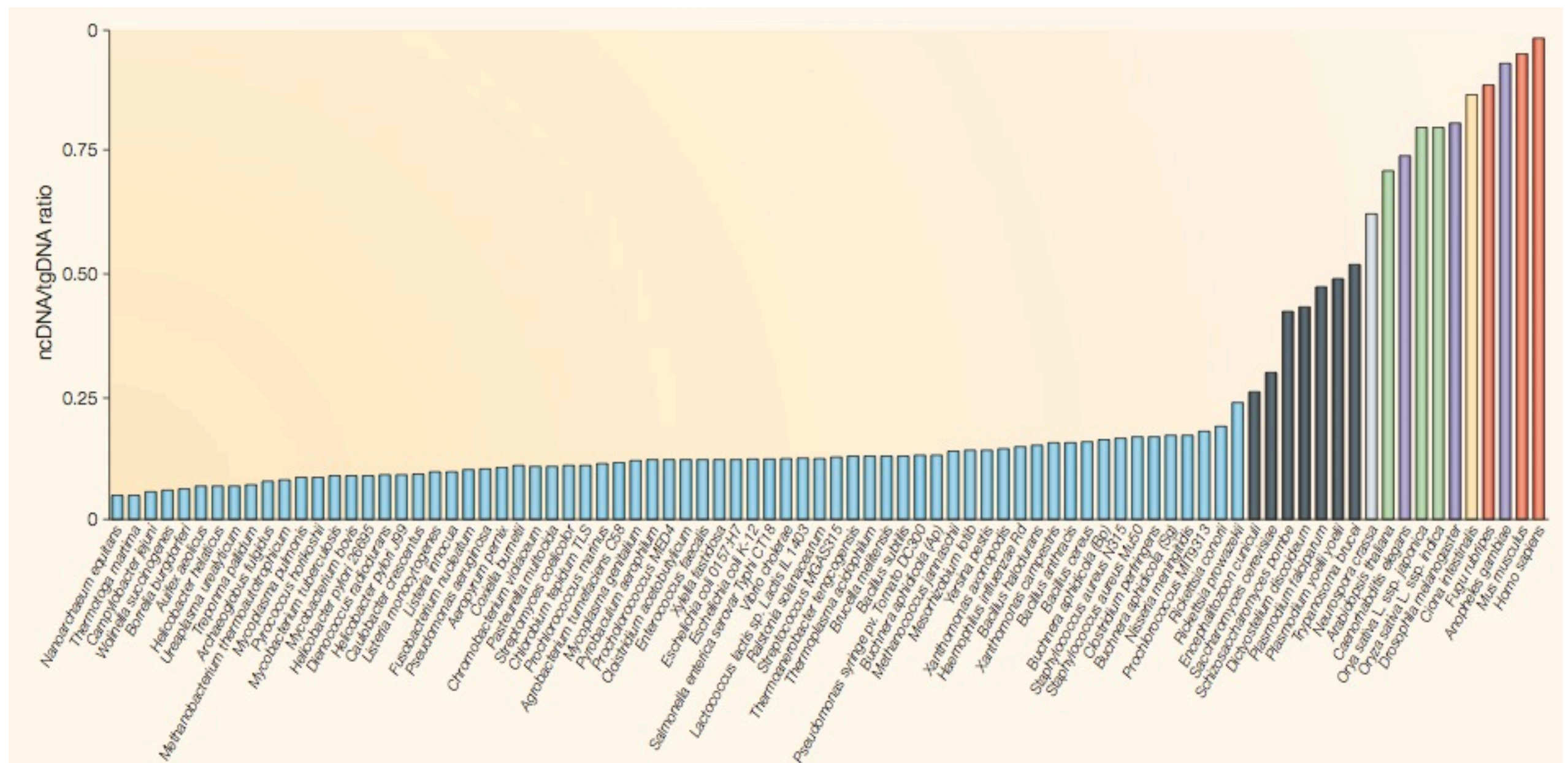


Figure 1 | The ratio of non-coding to protein-coding DNA rises as a function of developmental complexity. Prokaryotes have less than 25% non-coding DNA, simple eukaryotes have between 25 and 50% non-coding DNA and more complex fungi, plants and animals have more than 50%, rising to approximately 98.5% non-coding DNA in humans — which also have a genome size that is three orders of magnitude larger than prokaryotes. Note that this analysis corrects for ploidy, whereas pre-genomic estimations of the amount of DNA in different organisms did not. The different colours represent prokaryotes (bacteria and archaea) (blue), simple eukaryotes (black), *Neurospora crassa* (grey), plants (green), non-chordate invertebrates (nematodes, insects) (purple), *Ciona intestinalis* (urochordate) (yellow) and vertebrates (red). ncDNA, non-coding DNA; tgDNA, total genomic DNA. Reproduced with permission from REF. 77 © (2003) BioMed Central Ltd.

RNA regulation: a new genetics?

John S. Mattick

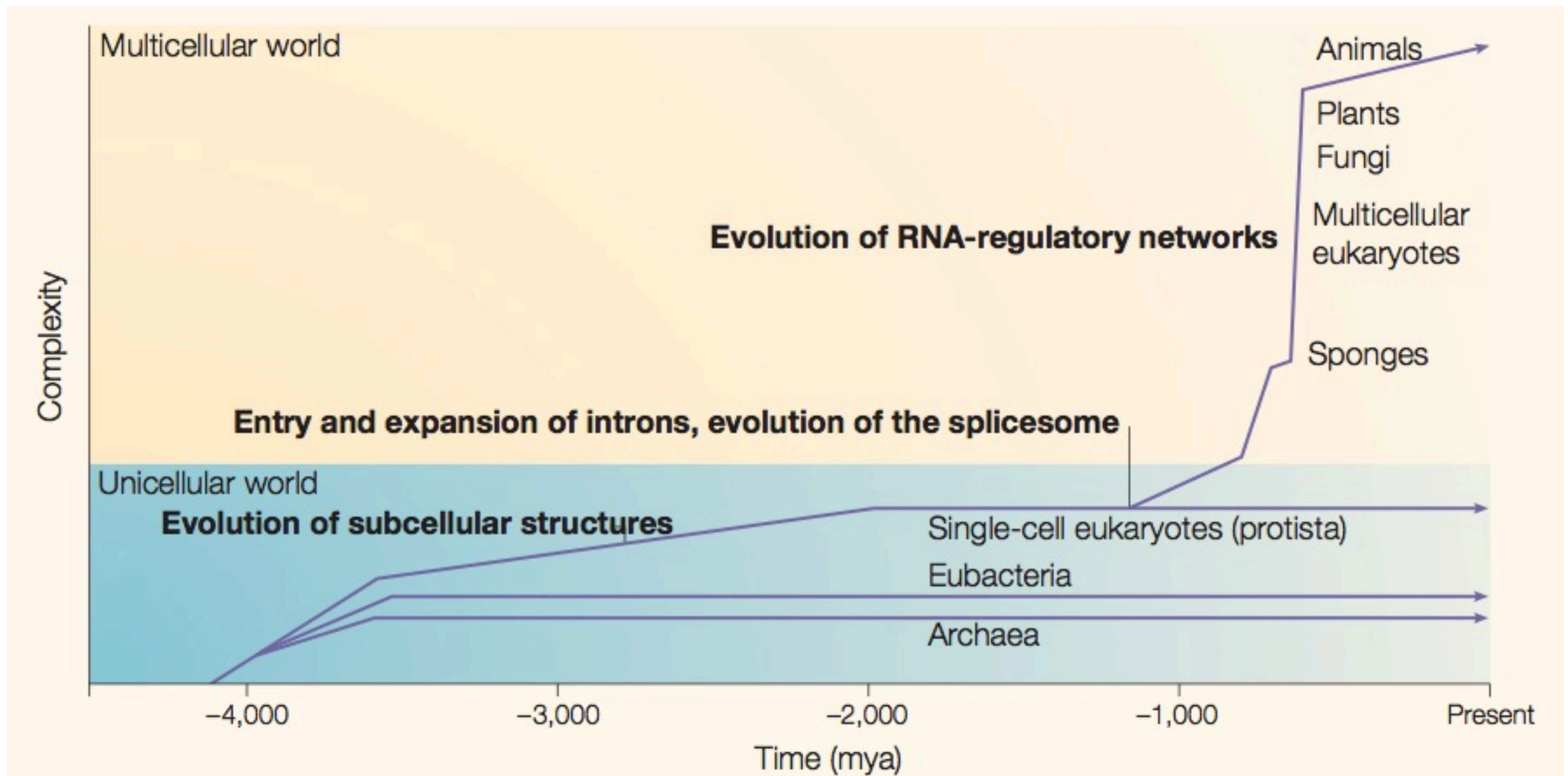
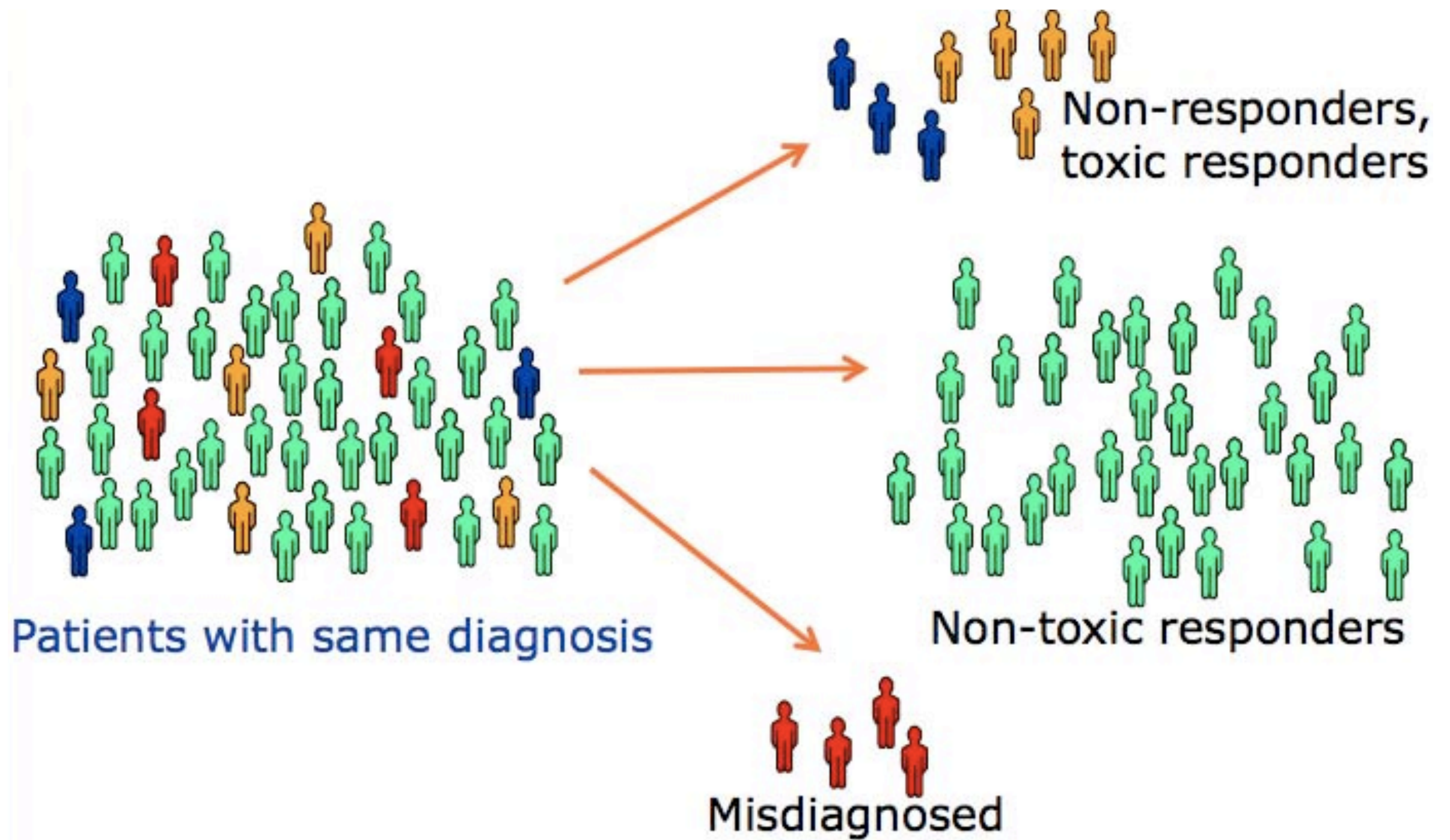


Figure 3 | **A simplified biological history of the Earth.** This graph is intended to present an overview.

RNA regulation: a new genetics?

John S. Mattick

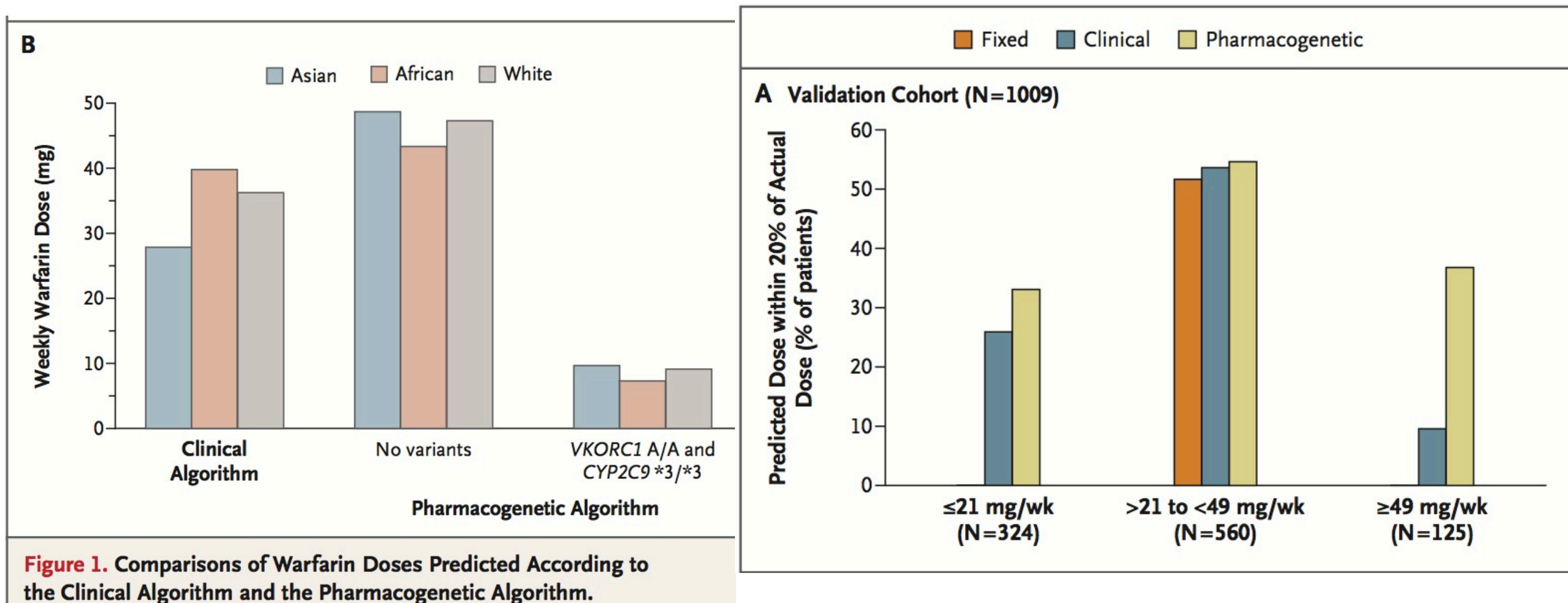


Estimation of the Warfarin Dose with Clinical and Pharmacogenetic Data

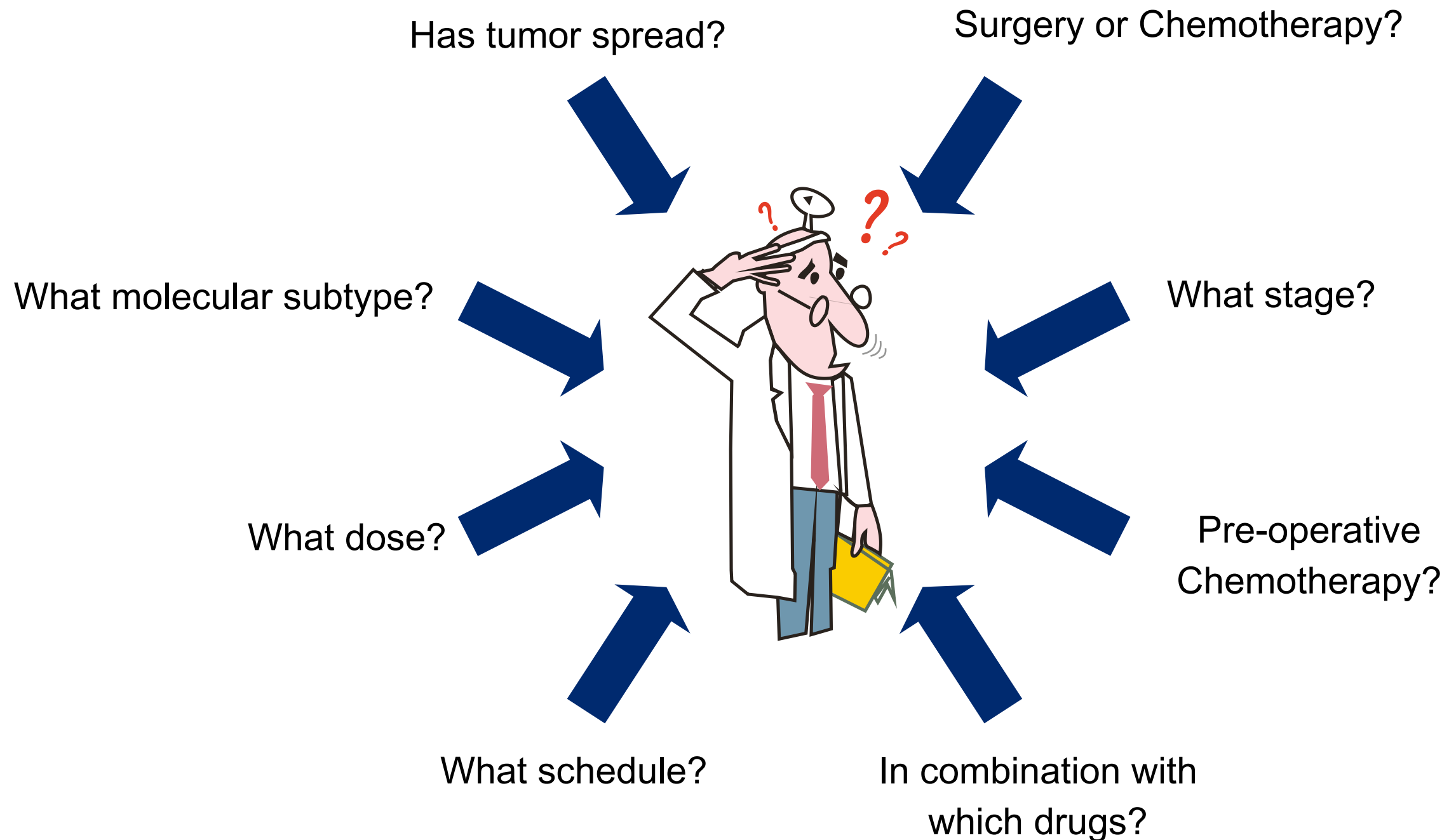
The International Warfarin Pharmacogenetics Consortium*

N Engl J Med 2009;360:753-64.

The investigators who performed the modeling and analysis did not have access to this validation set until after the final model was selected. A wide variety of numerical modeling methods were used for the data from the derivation cohort, including, but not limited to, support vector regression, regression trees, model trees, multivariate adaptive regression splines, least-angle regression, and Lasso, in addition to ordinary linear regression. Logarithmic and square-root transformations of doses were tested, in addition to a direct prediction of dose. Further details of the statistical modeling approaches that were tested and the evaluation methods that were used for selecting the best model from the derivation cohort are described in Section 3 in Supplementary Appendix 1.



Provider, Patient & Payor Faced With Bewildering Choices: The Current Practice of “Qualitative” Medicine



Using relative's DNA cracks crime, but privacy questions raised

By Jim Spellman, CNN

November 18, 2009 8:26 a.m. EST



Luis Jaimes-Tinajero was arrested after police tracked him down using his brother's DNA.

STORY HIGHLIGHTS

- Denver police this year tracked down a car theft suspect using his brother's DNA
- Investigators compared blood DNA from the scene to records in a criminal database
- DA says the method isn't for the courtroom, but for building leads in an investigation
- A defense attorney says it violates a person's "reasonable expectation of privacy"

Denver, Colorado (CNN) -- Using DNA to catch criminals has become common, but police in Denver, Colorado, this year demonstrated how the practice can be taken to a new level: They tracked down a suspect not through his DNA, but through that of his brother.

Here's how it happened, the Denver district attorney's office said:

In February 2008, two cars were broken into. Police found blood at both scenes and ran the

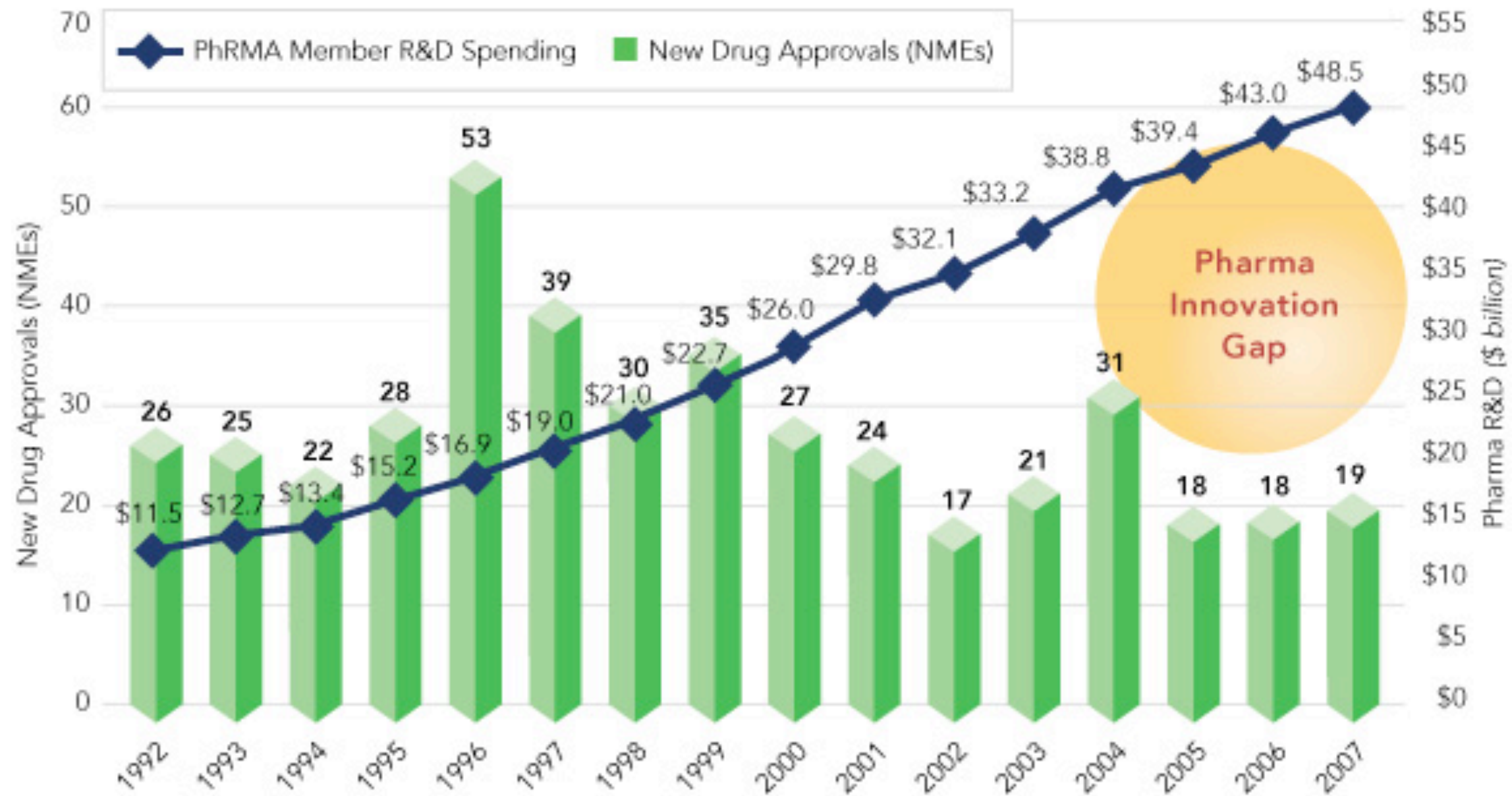
samples through DNA databases but couldn't find a match. Then, as part of a study being conducted by the district attorney's office, investigators used new software to see whether the DNA in the blood was close enough to potentially be from a family member of someone in the criminal DNA database.

The software came up with six potential matches. Five didn't pan out, but one led police to a convicted car thief and, ultimately, that man's brother, Luis Jaimes-Tinajero.

Jaimes-Tinajero pleaded guilty in September to one count of criminal trespass and received a sentence of two years' probation.

Innovation Gap Getting Wider: The Blockbuster Approach

INNOVATION GAP



Source: Burrill & Company, US Food and Drug Administration



UnitedHealth Settles Ingenix Suit: The Lawsuit Alleged That the Subsidiary Set Physician Payments Below the Doctors' Rate, Leaving Patients to Make Up the Difference.

Thursday, January 15, 2009 10:59 PM

(Source: Star Tribune, Minneapolis) By Chen May Yee, Star Tribune, Minneapolis

Jan. 16--UnitedHealth Group said Thursday it will pay \$350 million to customers and medical providers to settle a class-action lawsuit over the way it determined payments for out-of-network medical services going back nearly 15 years.

It is the biggest monetary settlement over medical payments by a single insurer, said the American Medical Association (AMA), which led the lawsuit.

"By using a flawed database to determine reimbursement rates for out-of-network care, insurers have increased profits at the expense of patients and physicians," AMA president Nancy Nielsen said in a statement. "By agreeing to the settlement, UnitedHealth Group has recognized the importance of restoring its relationship with patients and physicians by ending use of the rigged database."



ARRA Highlights

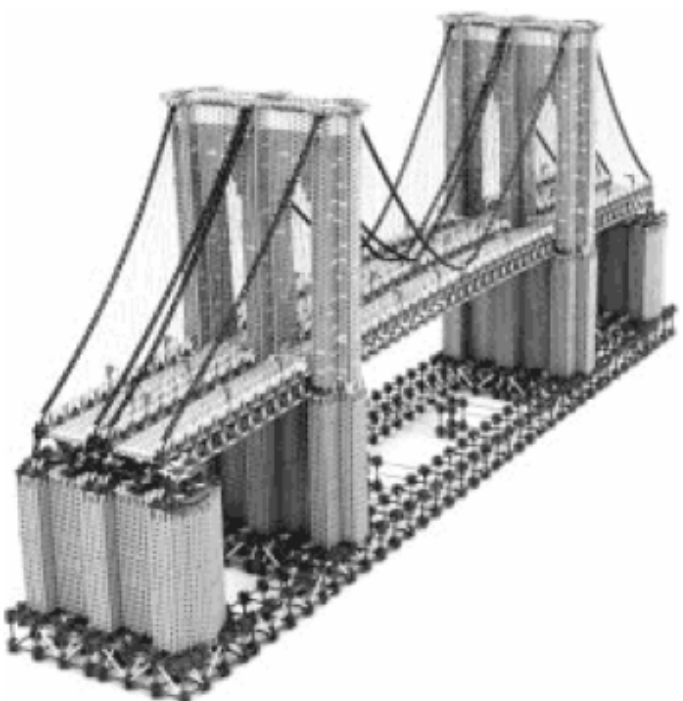
- Stimulus package provides over \$19 billion to support and promote the adoption of electronic health records (EHRs) for all Americans by 2014
- Health Information Technology for Economic and Clinical Health (HITECH) Act
 - New requirements, enforcement provisions, and penalties for covered entities, business associates, vendors, and others
- Codified the Office of the National Coordinator for Health Information Technology

Important characteristics

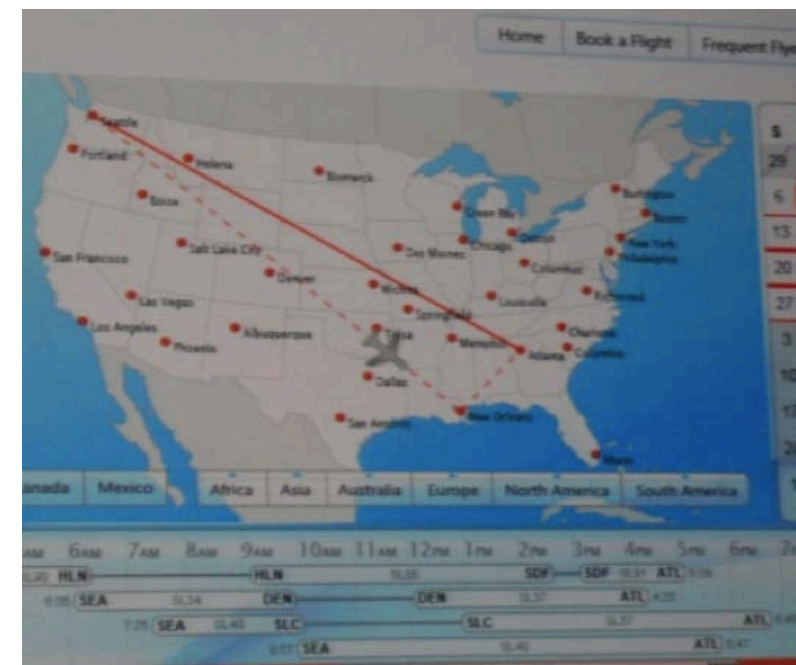
- We must integrate systems that may not have worked together before
- These are human systems, with differing goals, incentives, capabilities
- All components are dynamic—change is the norm, not the exception
- Processes are evolving rapidly too

Important characteristics

- We must integrate systems that may not have worked together before
- These are human systems, with differing goals, incentives, capabilities
- All components are dynamic—change is the norm, not the exception
- Processes are evolving rapidly too



We are not building something simple like a bridge or an airline reservation system



Healthcare is a **complex adaptive system**

A complex adaptive system is a collection of individual agents that have the freedom to act in ways that are not always predictable and whose actions are interconnected such that one agent's actions changes the context for other agents.

*Crossing the Quality Chasm, IOM,
2001; pp 312-13*

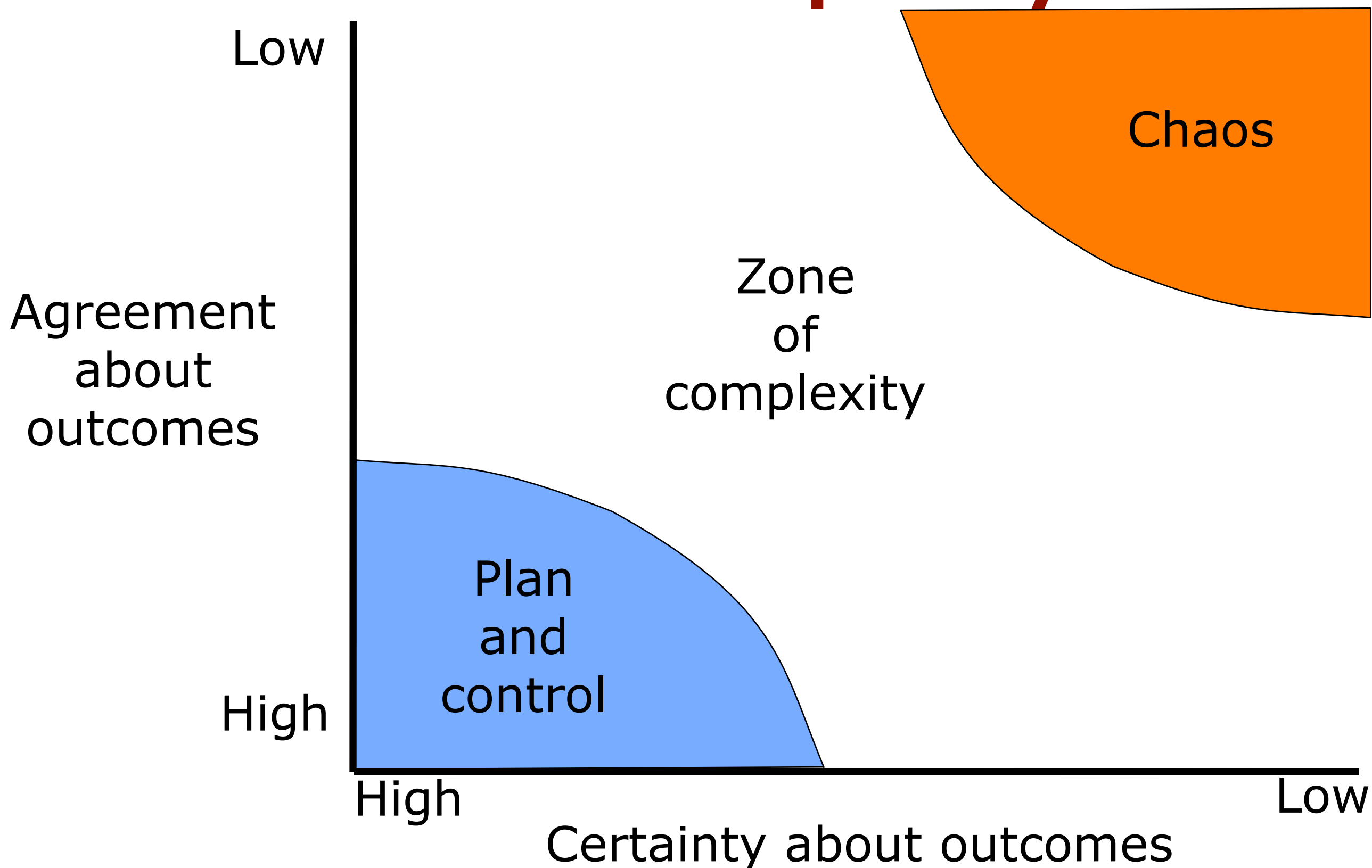
Healthcare is a **complex adaptive system**

A complex adaptive system is a collection of individual agents that have the freedom to act in ways that are not always predictable and whose actions are interconnected such that one agent's actions changes the context for other agents.

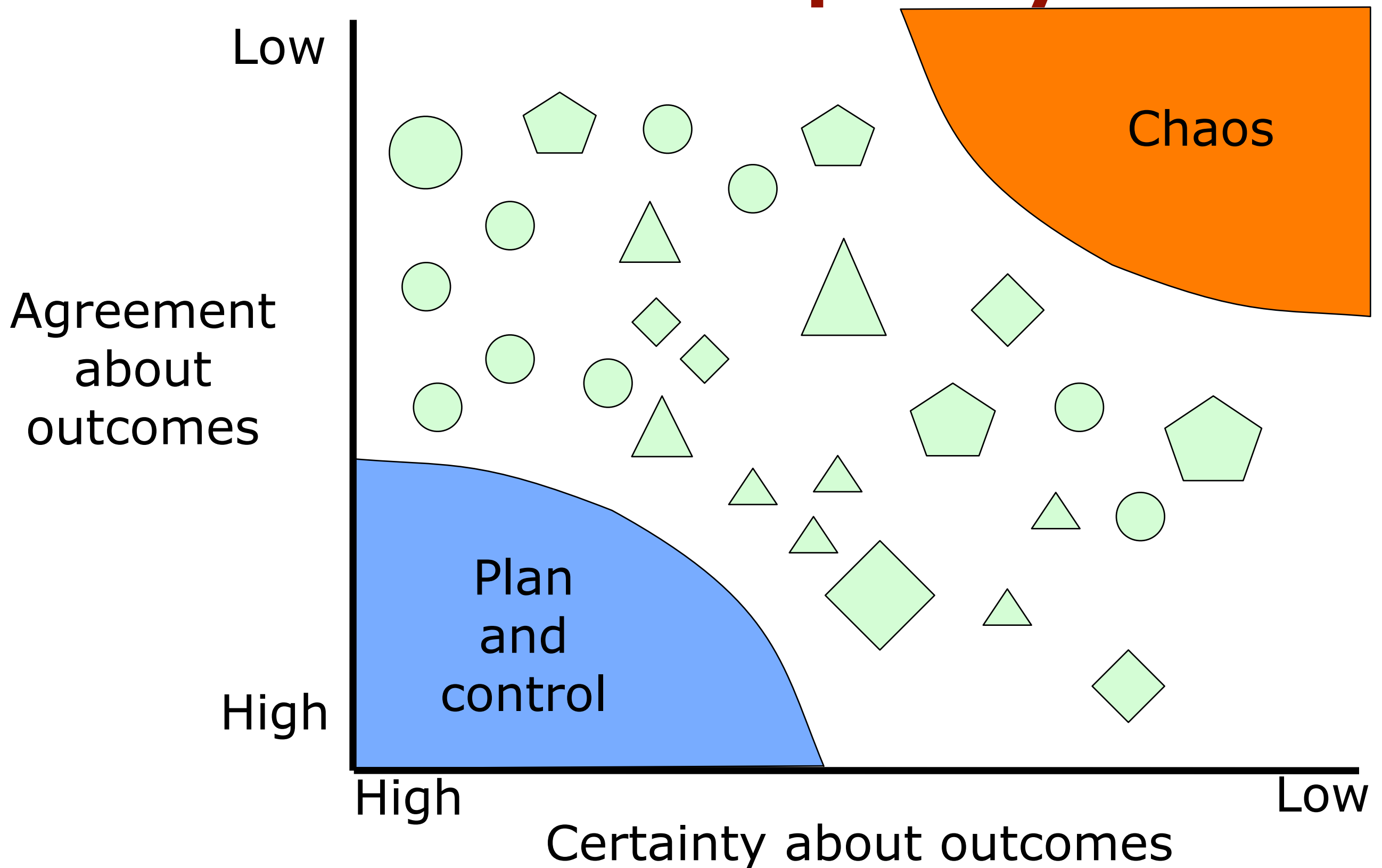
Crossing the Quality Chasm, IOM, 2001; pp 312-13

- Non-linear and dynamic
- Agents are independent and intelligent
- Goals and behaviors often in conflict
- Self-organization through adaptation and learning
- No single point(s) of control
- Hierarchical decomposition has limited value

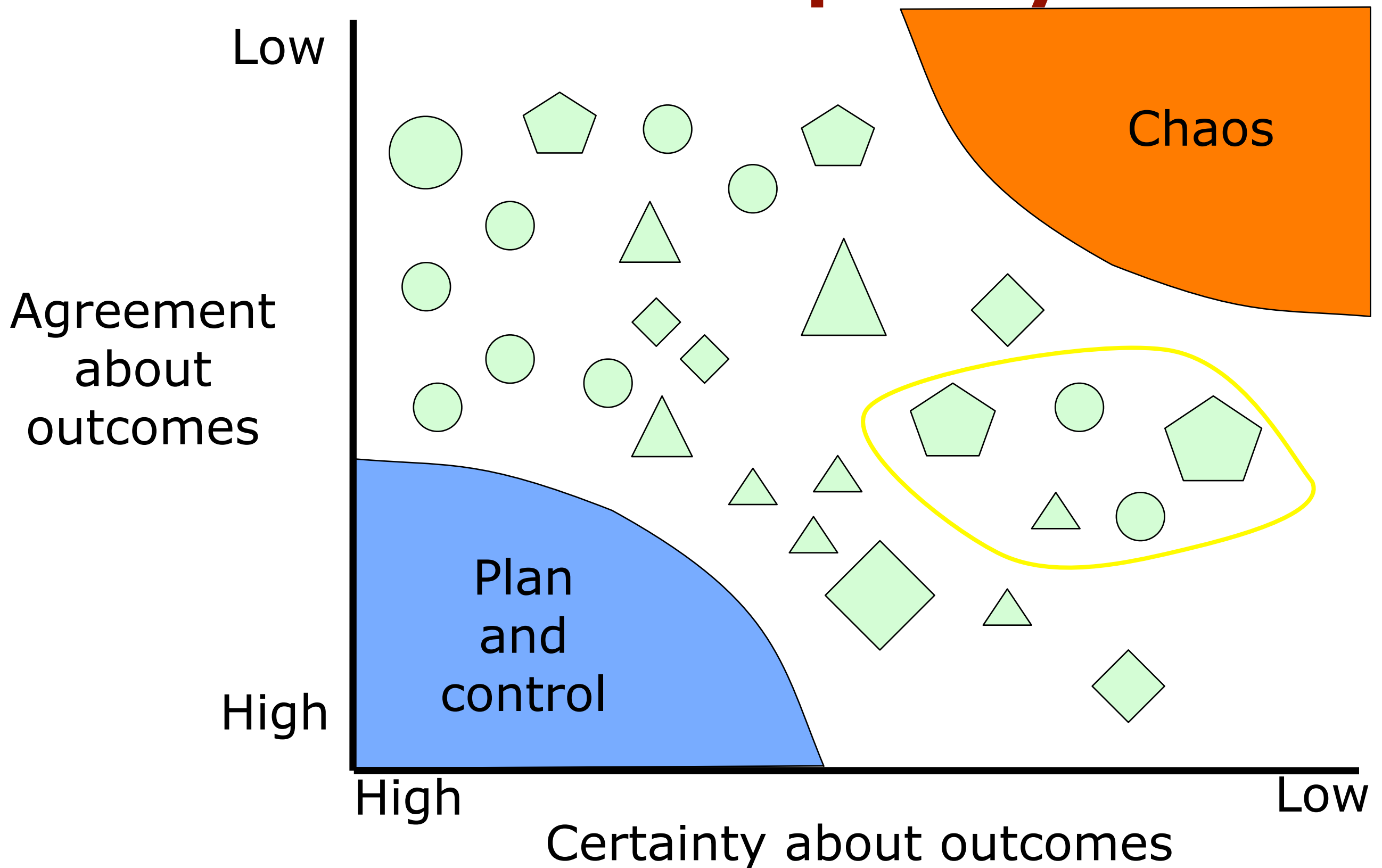
We need to function in the **zone of complexity**



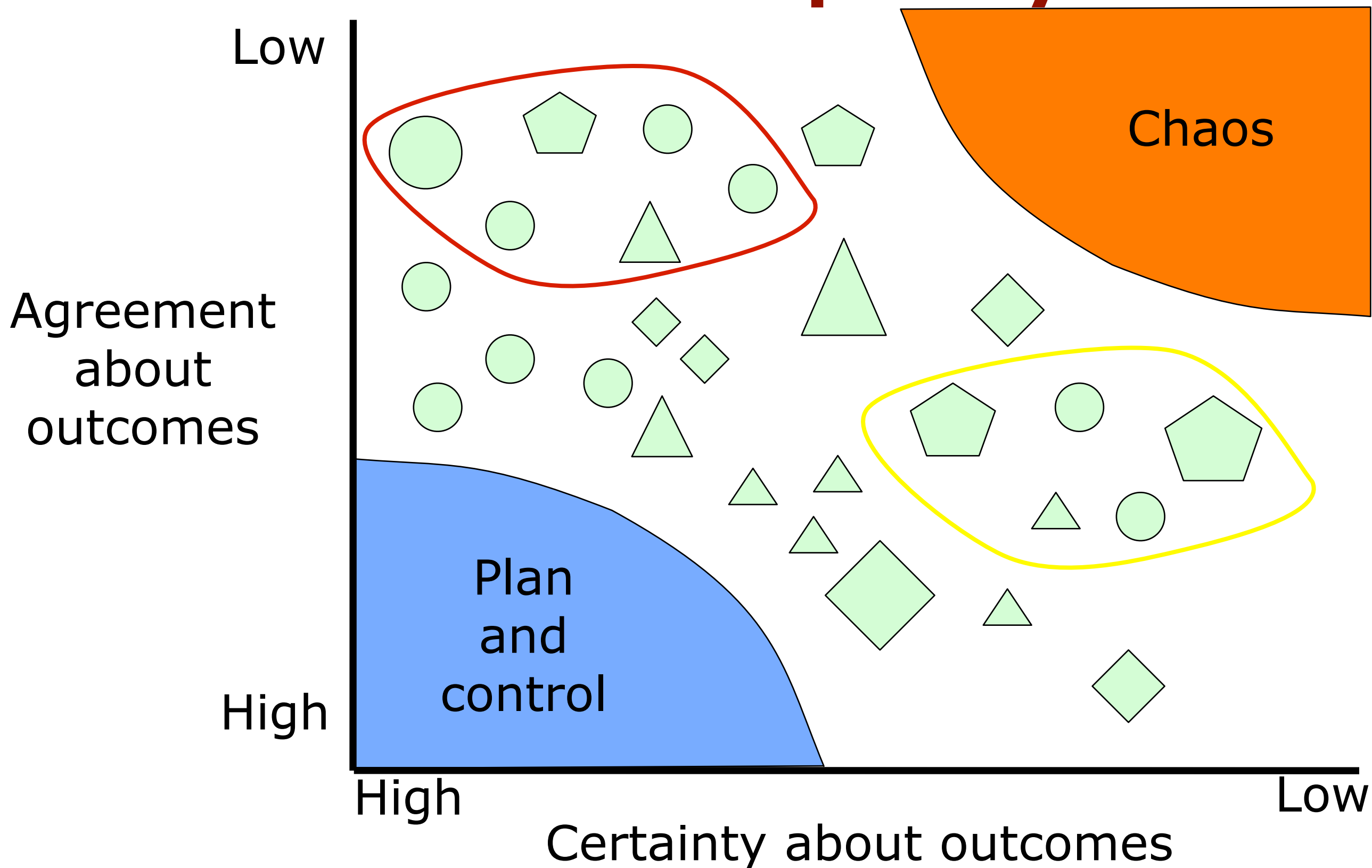
We need to function in the **zone of complexity**



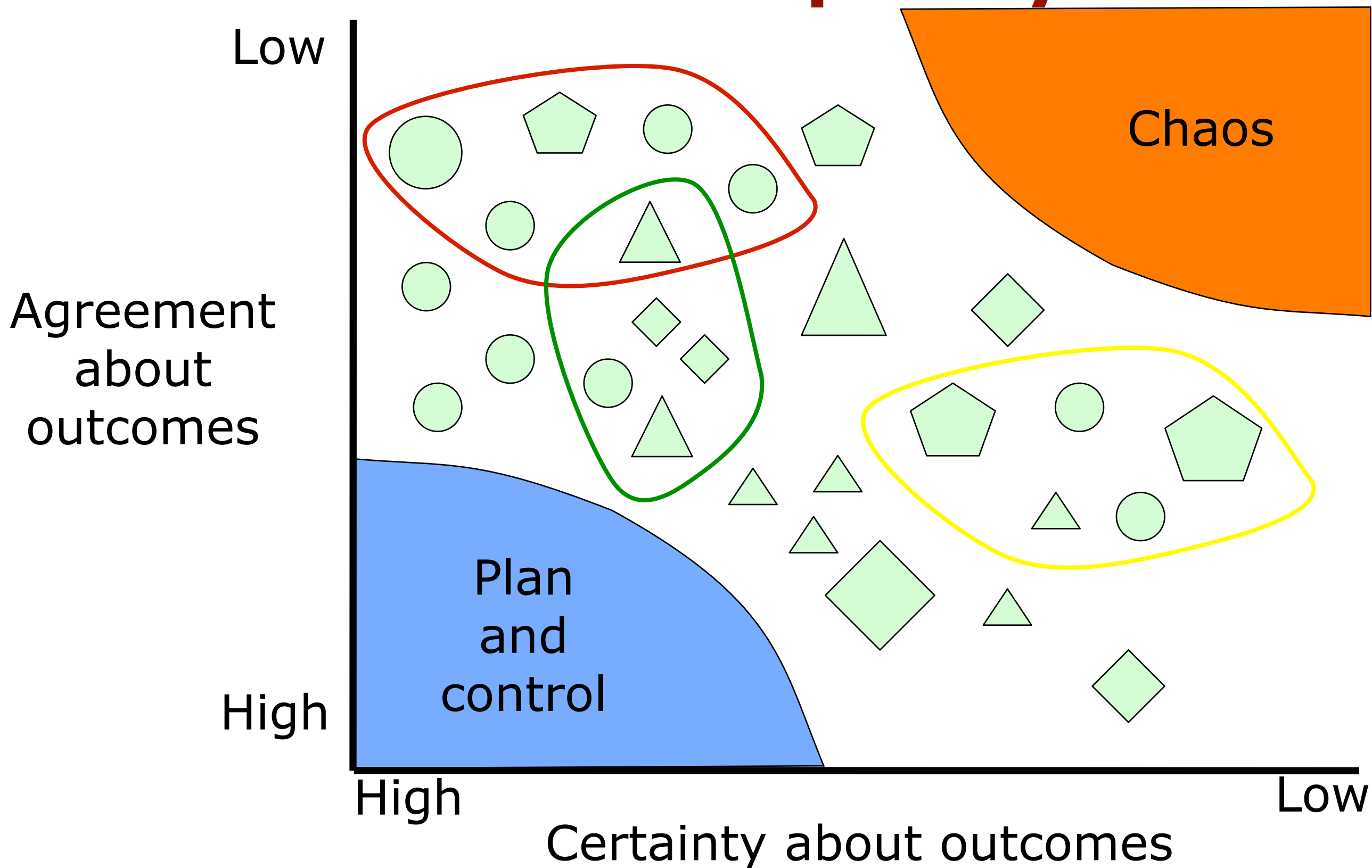
We need to function in the **zone of complexity**

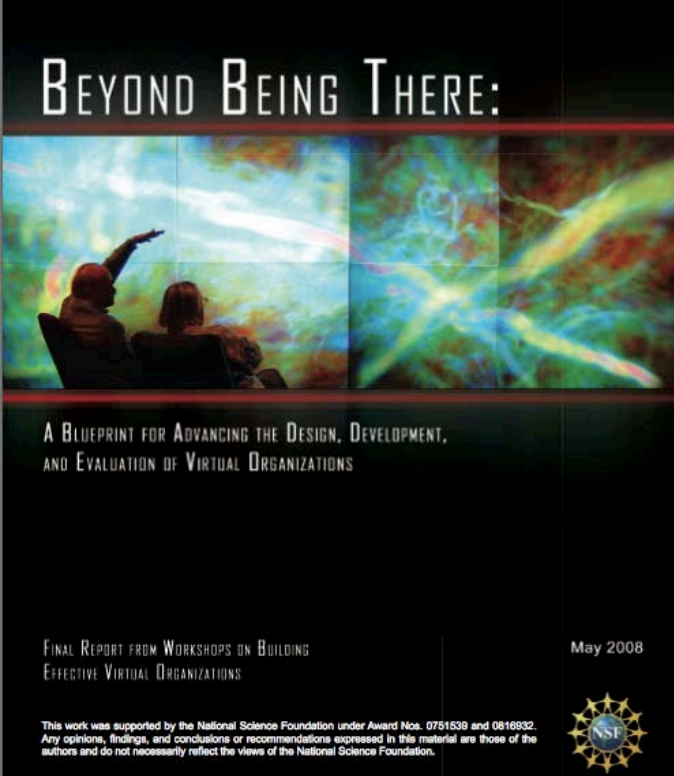


We need to function in the **zone of complexity**



We need to function in the **zone of complexity**





We call these groupings **virtual organizations (VOs)**

A set of individuals and/or institutions engaged in the controlled sharing of resources in pursuit of a common goal

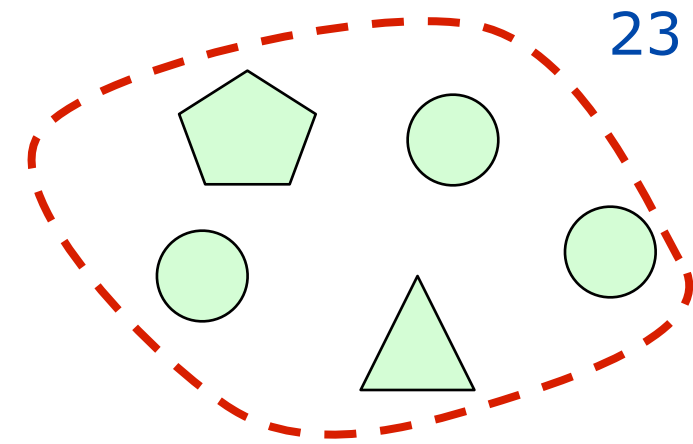
Healthcare = dynamic, overlapping VOs, linking

- ◆ Patient – primary care
- ◆ Sub-specialist – hospital
- ◆ Pharmacy – laboratory
- ◆ Insurer – ...

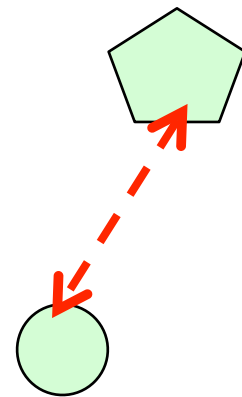
But U.S. health system is marked by fragmented and inefficient VOs with insufficient mechanisms for controlled sharing

I advocate ... a model of virtual integration rather than true vertical integration ... G. Halvorson, CEO Kaiser

The Grid paradigm



- Principles and mechanisms for dynamic VOs
- Leverage service oriented architecture (SOA)
- Loose coupling of data and services
- Open software, architecture



Computer
science



Physics



Astronomy

Engineering



Biomedicine



Healthcare

Biology



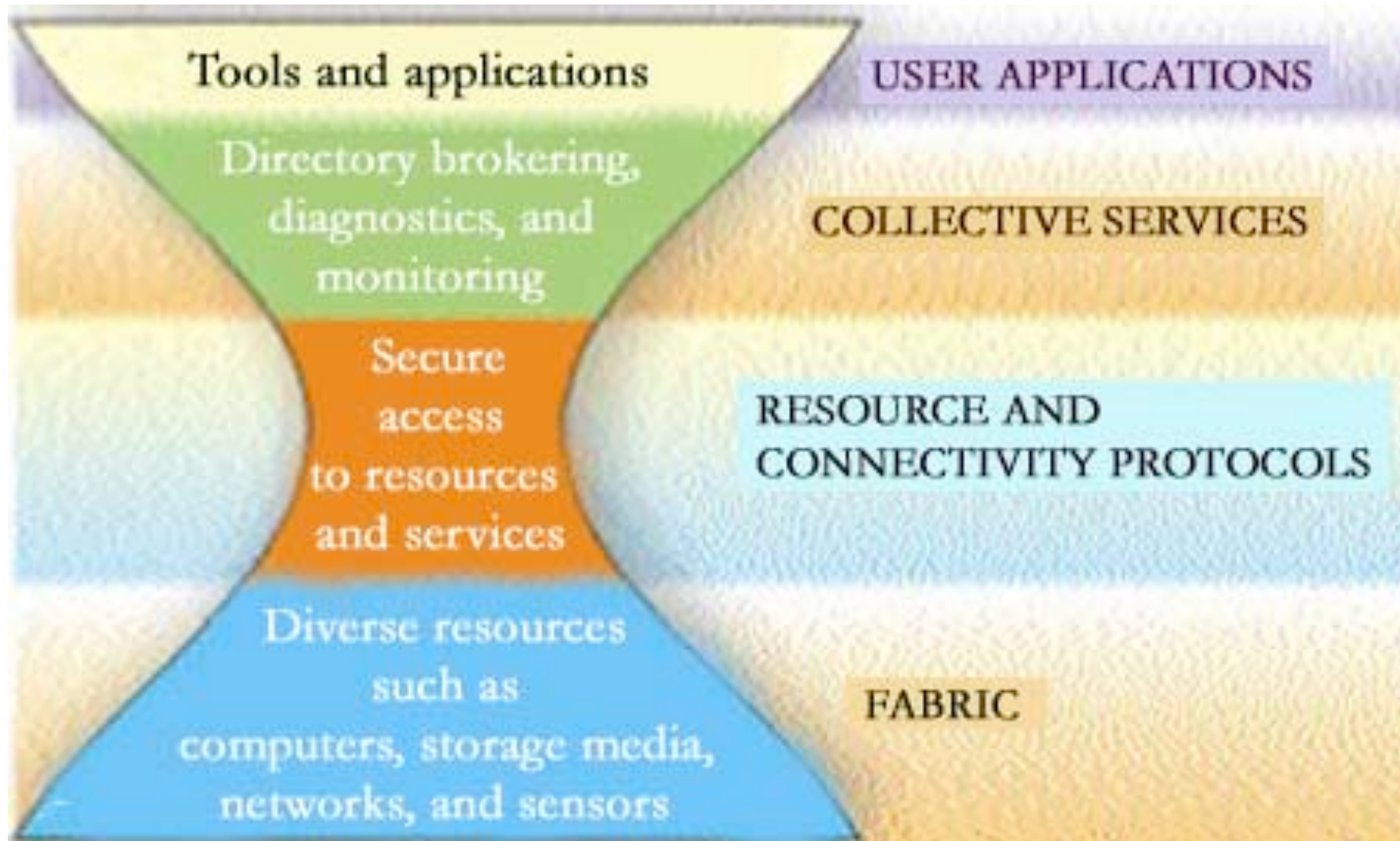
1995

2000

2005

2010

Hour Glass Model



Grid Characteristics

- Multi-organizational
- Securibility
- Flexible policy/sociability
- Extensibility
- Redundancy
- Robust in multiple industries/stability
- Needs no central ownership/managability, etc...



Globus Toolkit

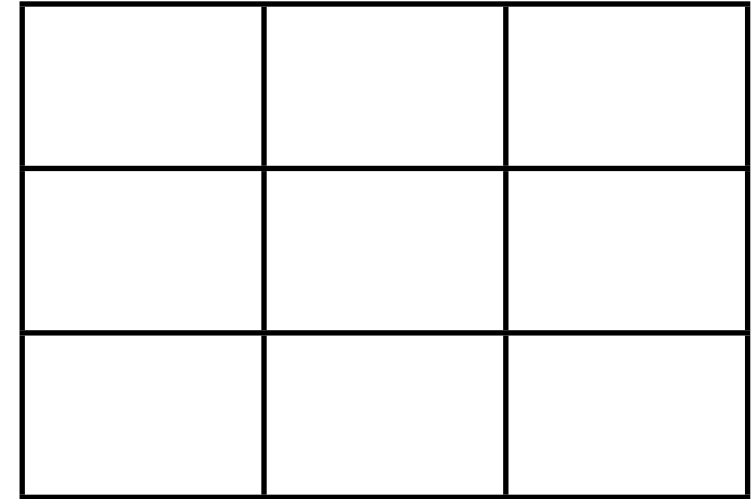
Assortment of Components for Grid Builders

- Focus on Connectivity and Resource layers
 - ◆ GRAM, GSI-OpenSSH: Run programs
 - ◆ GridFTP: Access file systems
 - ◆ OGSA-DAI, caGrid: Access databases
 - ◆ GSI, Myproxy, GAARDS: Security
 - ◆ XIO, Java Core, C Core: Communication
- A few simple Collective layer components
 - ◆ RLS: Replica tracking
 - ◆ RFT: Reliable file transfer

Hosted Federation Services

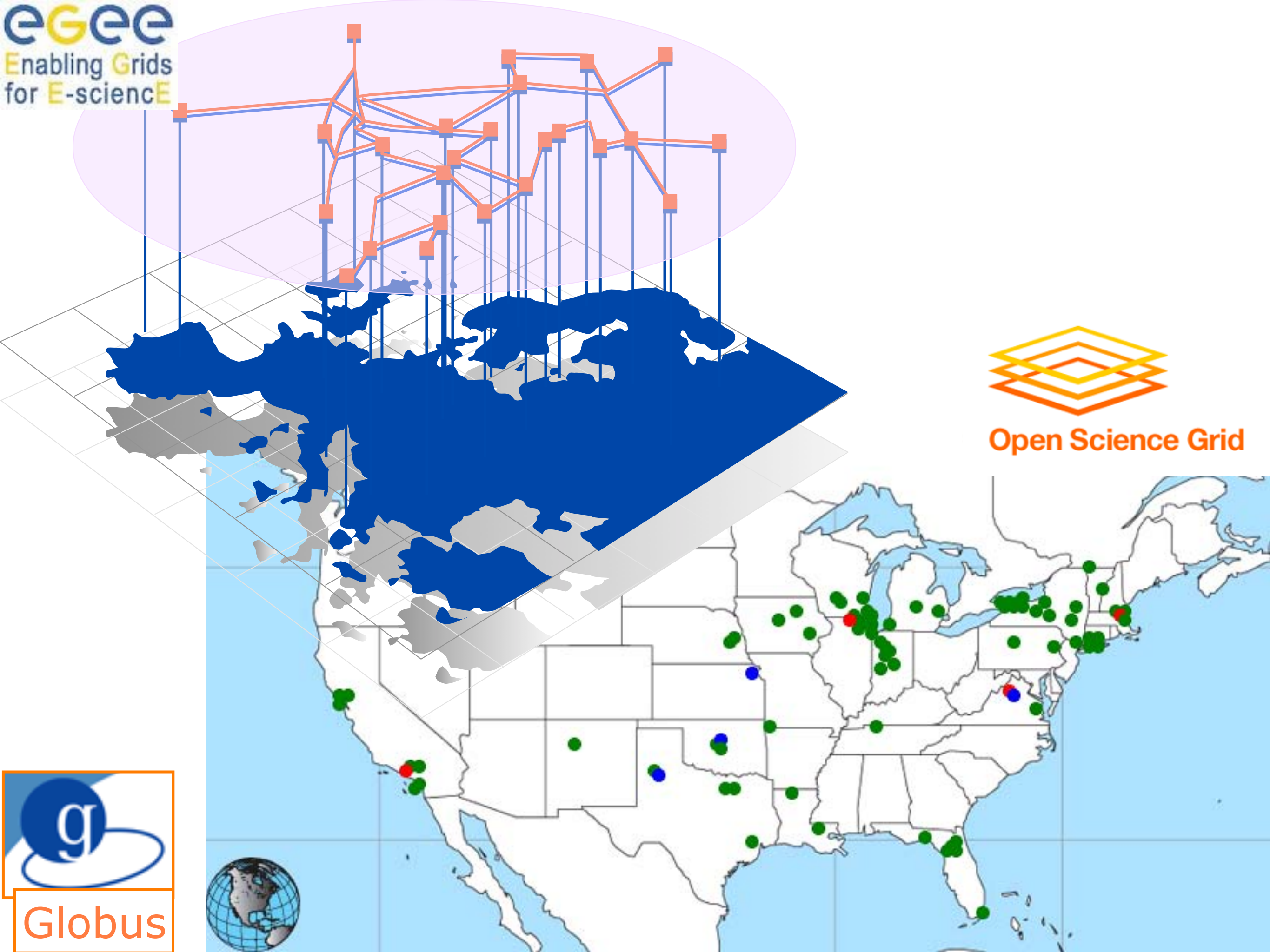


Emerging from virtualization
of machines, business
models, flexible capacity



Evolving from virtualization
of organizations, social
models, flexible capabilities,
security, open services...

Infrastructure on demand



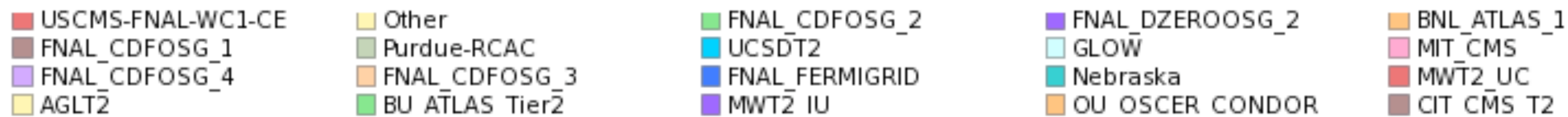
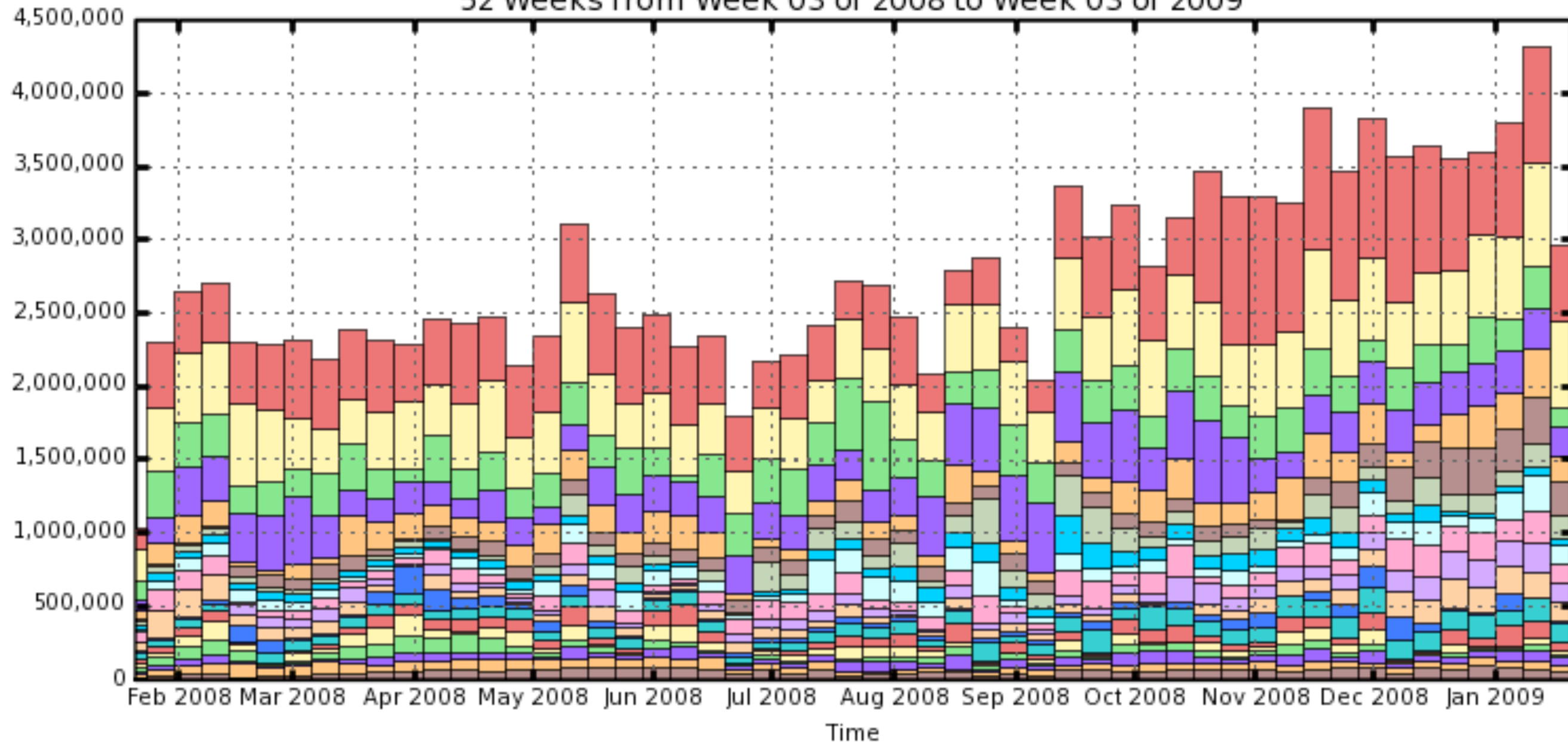
Open Science Grid



Globus

Hours Spent on Jobs By Facility

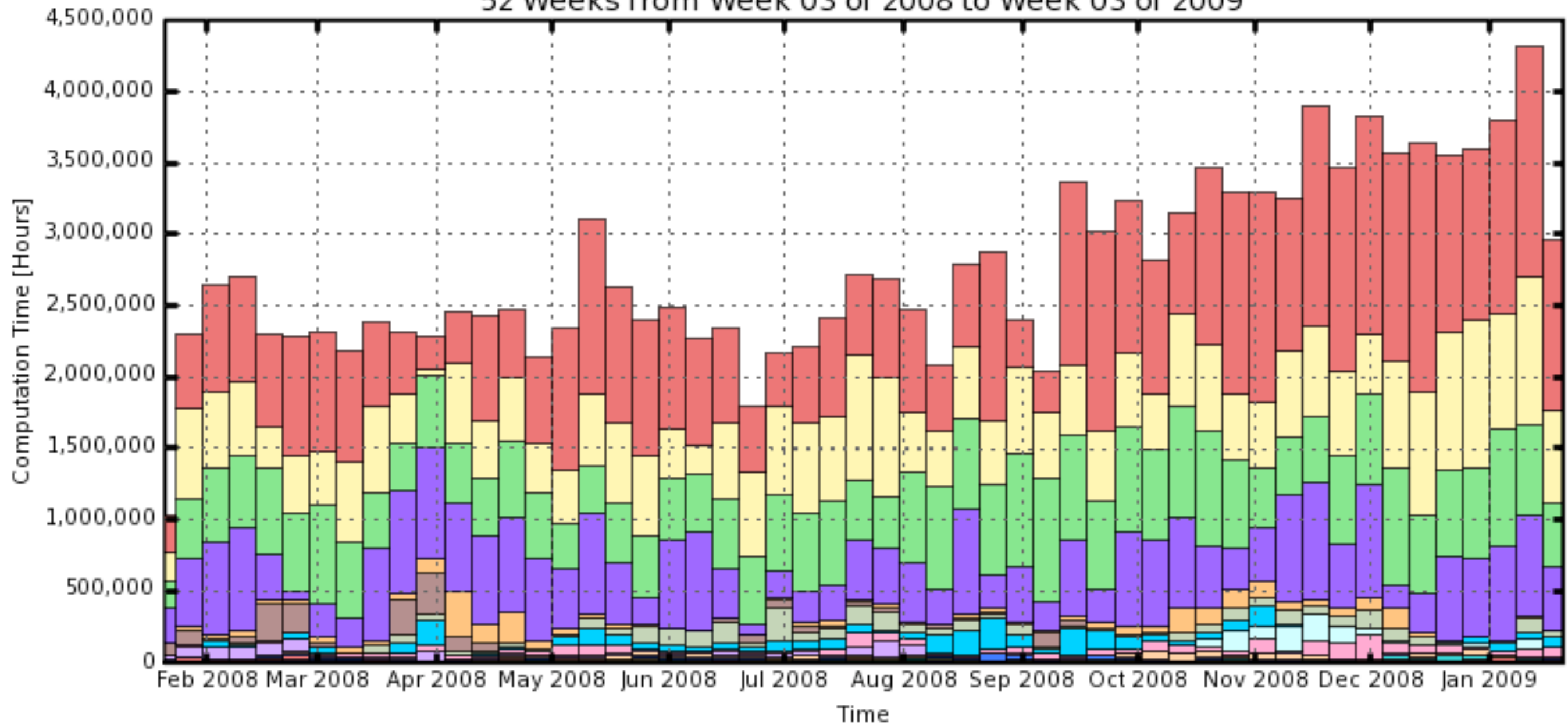
52 Weeks from Week 03 of 2008 to Week 03 of 2009



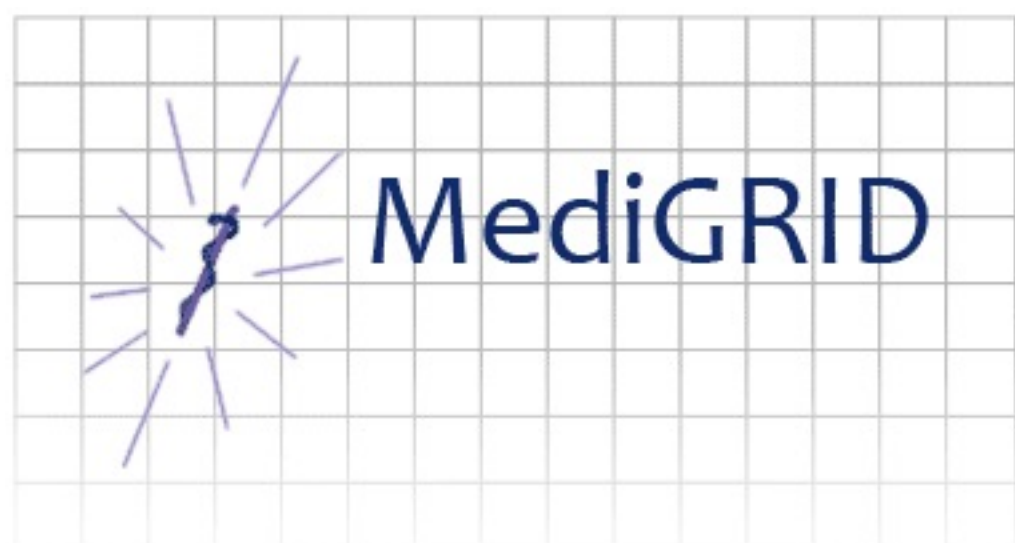
Maximum: 4,312,733 , Minimum: 1,024,283 , Average: 2,755,248 , Current: 2,966,011

Hours Spent on Jobs By VO

52 Weeks from Week 03 of 2008 to Week 03 of 2009



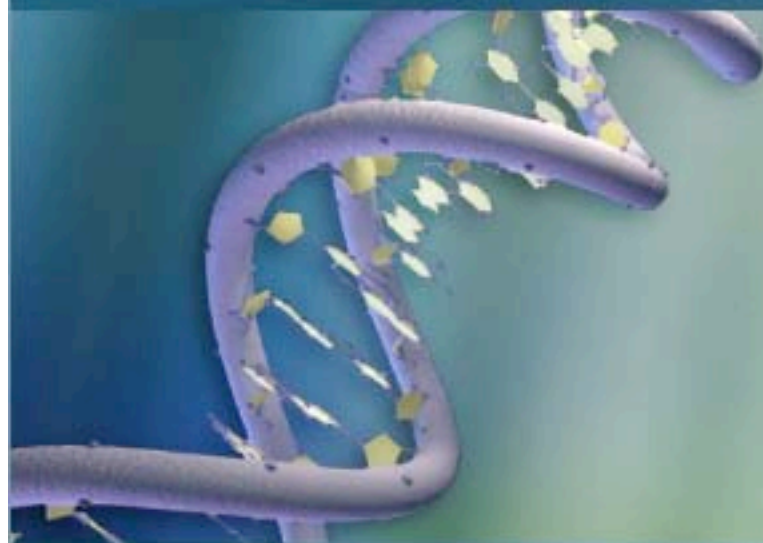
Maximum: 4,312,733 Hours, Minimum: 1,024,283 Hours, Average: 2,755,248 Hours, Current: 2,966,011 Hours



WISDOM

Initiative for grid-enabled drug discovery against neglected and emergent diseases

WISDOM initiative aims to demonstrate the relevance and the impact of the grid approach to address drug discovery for neglected and emergent diseases. It gathers several partners around the world conscious of the urgency of working on these diseases. They are using grid infrastructure to organise and accelerate their research. They deploy production experiment of virtual screening at a large scale against diseases, called a data challenge. The acronym WISDOM, for Wide In Silico Docking On Malaria, comes from the first screening experiment at a large scale against malaria and is now used as a generic name for the initiative.



The Conduit for Biomedical Research

Enter search keyword



BIRN Wiki



BIRN Launches New Website Highlighting Fresh Priorities October 13, 2009

The Biomedical Informatics Research Network (BIRN) today launched a completely new website (www.birncommunity.org) that highlights its broadened focus, including services for the entire biomedical community [...]

CardioVascular Research Grid partners with BIRN October 6, 2009

The CardioVascular Research Grid (CVRG) and Biomedical Informatics Research Network (BIRN) are partnering to create an infrastructure for sharing cardiovascular data and data analysis tools, [...]

[More News](#)

Welcome to cagrid.org



What is caGrid? caGrid is open source grid software infrastructure aimed at enabling multi-institutional data sharing and analysis. caGrid supports a wide range of use cases in basic, translational, and clinical research. Key innovations in caGrid support large scale, secure, semantically meaningful data sharing and analysis. caGrid also provides enterprise tools and services for the development and management of caGrid services and applications. The most recent version of caGrid, version 1.3, represents the sixth major release of caGrid since early 2005. [Learn more...](#)

About this site. cagrid.org has a navigation bar with buttons for six sections. [Learn more...](#)

Recent News

Title	Author	Date Posted
 Call for Arch and VCDE F2F Birds of a Feather Sessions	Justin Permar	Apr 21, 2009
 May User Group Call	William Stephens	Apr 16, 2009
 caGrid 1.3 Released	Scott Oster	Mar 31, 2009
 April User Group Call	William Stephens	Mar 26, 2009
 caBIG Production Grid Maintenance for caGrid 1.3 Deployment	Justin Permar	Mar 25, 2009
 User Group Call	William Stephens	Mar 06, 2009
 caGrid 1.3 Training Grid Registration Issues	William Stephens	Mar 06, 2009

The Team

The caGrid "core" development team for the official caGrid release is composed of developers from many organizations. Learn more about [the development team](#) and [the knowledge center team](#).

Project of the Month



Data services provide the ability to share data on the Grid. [Learn more...](#)



CardioVascular Research Grid

Navigation

- Organization
- Research
- CVRG Portal
- CVRG Wiki
- Presentations
- Press
- CVRG Collaboration Projects
- Other Resources

The aim of the CardioVascular Research Grid (CVRG) project is to create an infrastructure for sharing cardiovascular data and data analysis tools. The CVRG will support national and international multi-institutional collaborations in cardiovascular science, and will build on and extend tools developed in the [caBIG](#) and [BIRN](#) projects. The CVRG project is supported by the [National Heart Lung & Blood Institute](#). The project is based at the [Institute for Computational Medicine](#) at Johns Hopkins University, in collaboration with the [Center for Comprehensive Informatics](#) at Emory University.

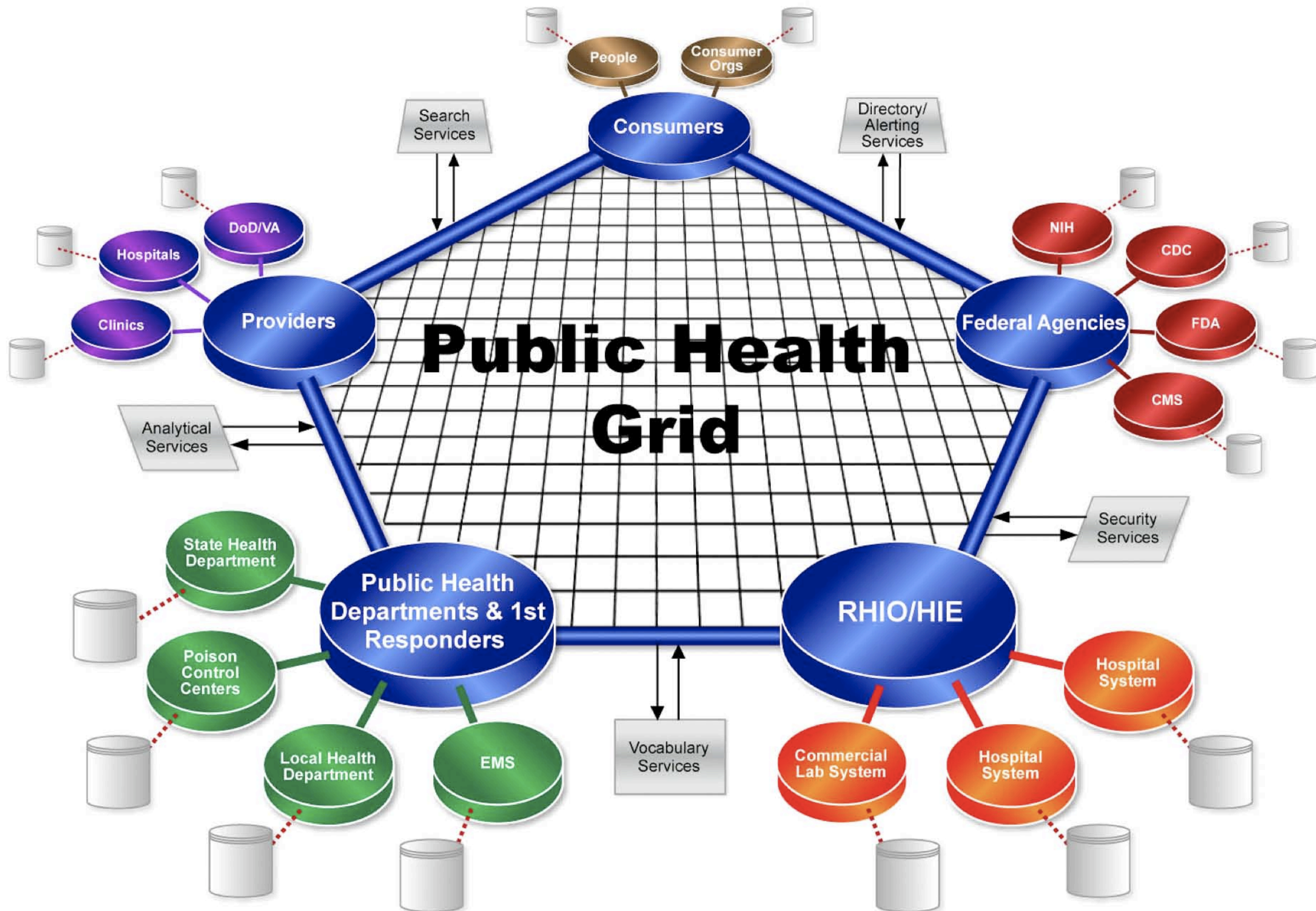
Within the navigation bar, hold your mouse pointer over category links for a summary of the content.

In the News

CardioVascular Research Grid partners with BIRN

October 7, 2009

The [CardioVascular Research Grid \(CVRG\)](#) and [Biomedical Informatics Research Network \(BIRN\)](#) are partnering to create an infrastructure for sharing cardiovascular data and data analysis tools, both organizations announced today.



CTSA Committees & Activities

[Consortium Committees](#)

[CTSA Consortium Steering Committee](#)

[CTSA Consortium Child Health Oversight Committee](#)

[Strategic Goal Committees](#)

[National Clinical and Translational Research Capability](#)

[Training & Career Development of Clinical/Translational Scientists](#)

[Enhancing Consortium-Wide Collaborations](#)

[Enhancing the Health of Our Communities and the Nation](#)

[T1 Translational Research](#)

[Key Function Committees](#)

[Administration](#)

[Biostatistics / Epidemiology / Research Design](#)

[Clinical Research Ethics](#)

Participating Institutions

The University of Chicago Institute for Translational Medicine

Chicago, IL

[Visit Site](#)

Principal Investigator: [Julian Solway, MD](#)

Participating Institutions:

- The University of Chicago, Chicago, IL
- Argonne National Laboratory (ANL), Argonne, IL
- Illinois Institute of Technology (IIT), Chicago, IL
- University of Chicago Medical Center, Chicago, IL
- Access Community Health Network, Chicago, IL
- Advocate Health Care, Chicago, IL
- Additional University of Chicago Participating Divisions
 - Biological Sciences Division
 - Physical Sciences Division
 - Division of Social Sciences
 - School of Social Service Administration
- Additional University of Chicago Centers
 - Center for Health and the Social Sciences
 - Institute for Molecular Pediatric Science
 - MacLean Center for Clinical Medical Ethics
 - University of Chicago Cancer Research Center
 - Institute for Genomics and Systems Biology
 - Institute for Biophysical Dynamics
 - Center for Interdisciplinary Health Disparities Research
 - Computation Institute

Quick Links

[Federated Wiki Access](#)

[Resources for Researchers](#)

[Building Connections](#)

[Calendar](#)

[e-Newsletter](#)

[Communication Toolkit](#)

[Governance Manual](#)

[Institution Search](#)

[CTSA Committee Participants](#)

Attribute-based Access Control (ABAC)

Generalization of attributes

Role-based Access Control (RBAC)

Separation of role/group from rule admin

Identity-based authZ

Most simple - not scalable

Mandatory Access Control (MAC)

Clearance, classification, compartmentalization

POSIX ACLs/MS-ACLs

Finer-grained admin policy

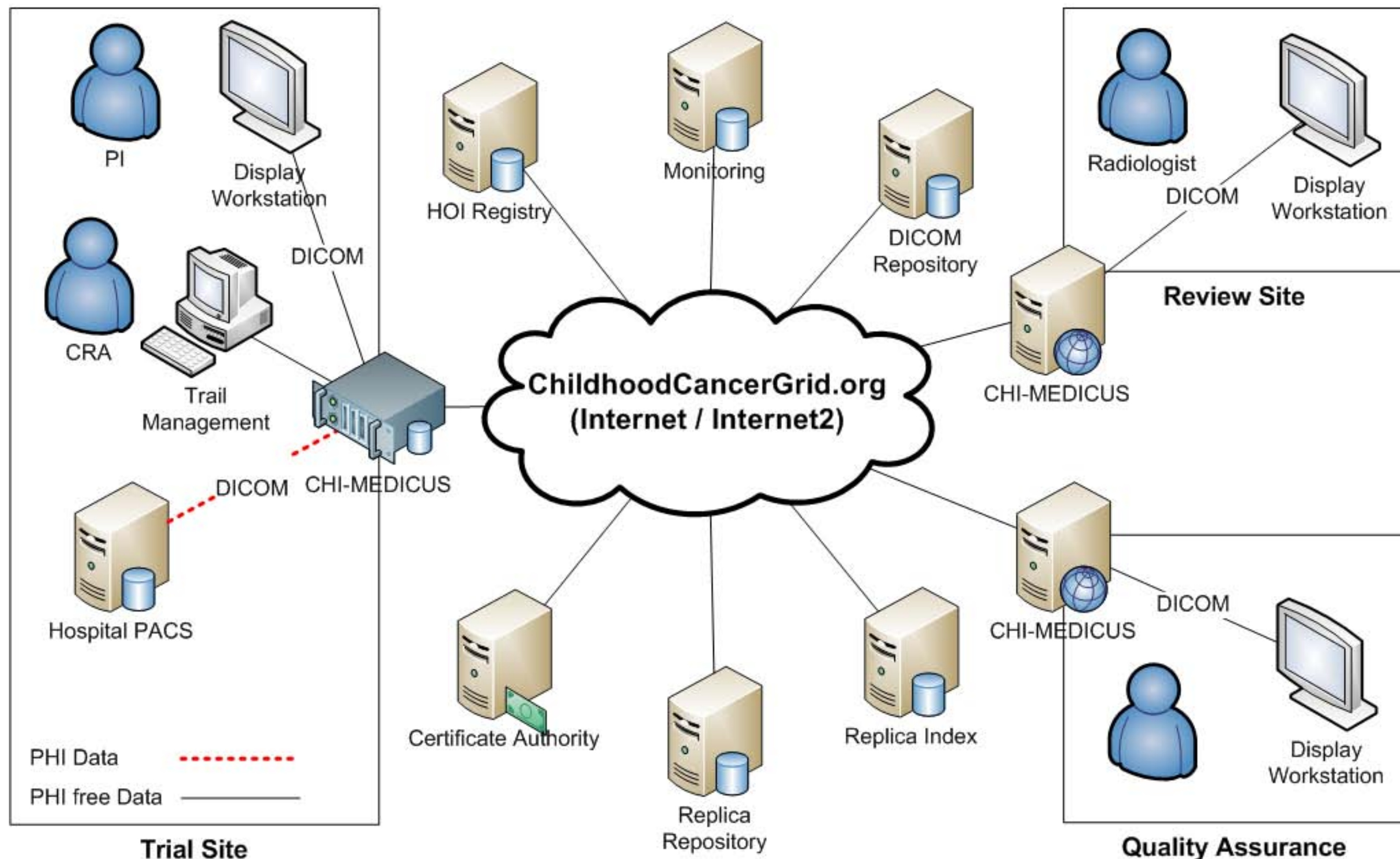
Unix Access Control Lists (Discretionary Access Control: DAC)

Groups, directories, simple admin

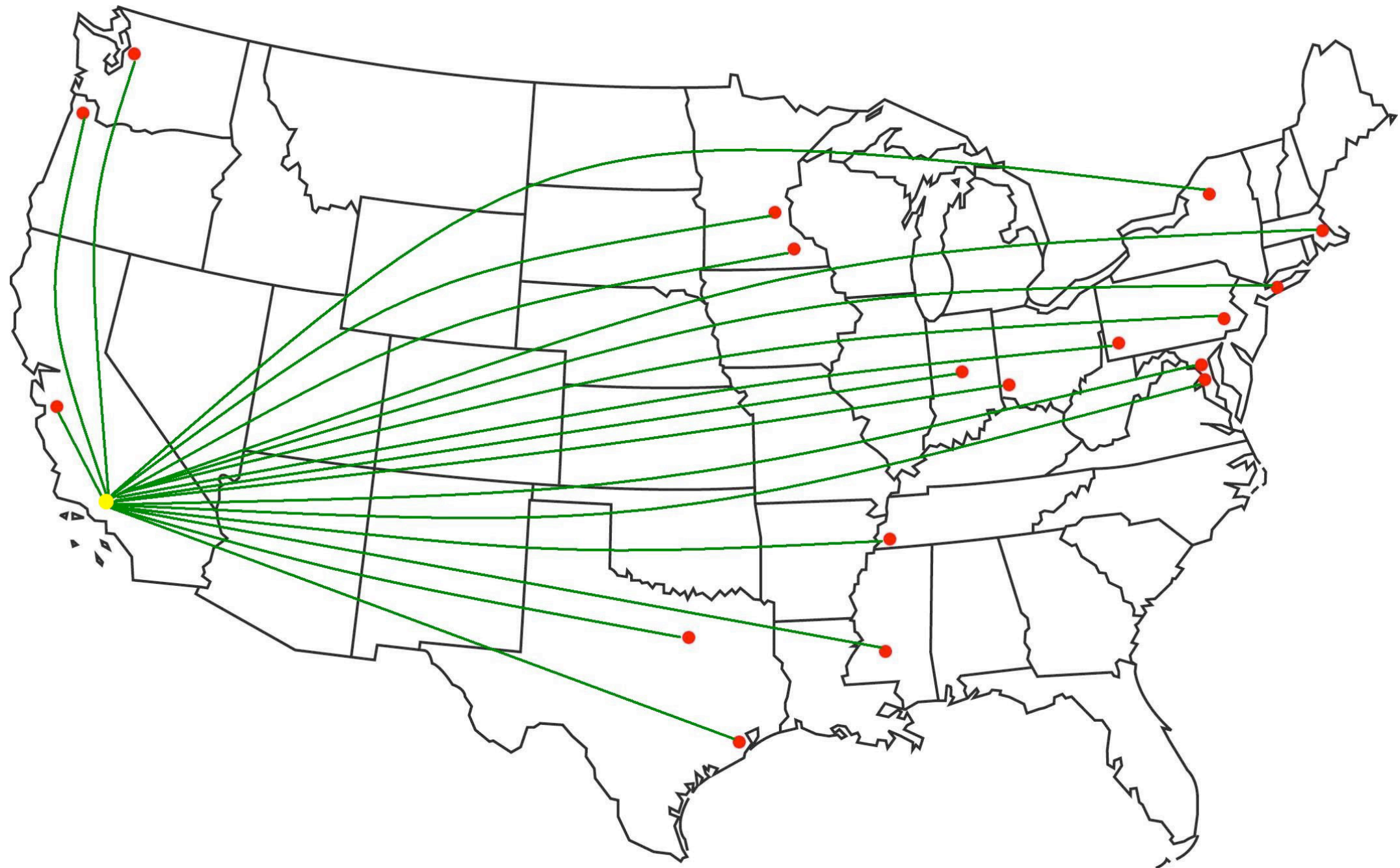
>>> Policy language abstraction level and expressiveness >>>

Data movement in clinical trials

Multi Center Clinical Cancer Trials Image Capture & Review Infrastructure



Children's Oncology Grid



Semantic expressivity is generally problematic in biomedical data

- Biomedical concepts are context dependent
 - ◆ For billing data, ICD and CPT works
 - ◆ For quality/effectiveness/research more detail is required
- Encode data for semantic interoperability and re-use— or collect specific to context?
 - ◆ Physicians prefer free text
 - ◆ Biomedical researchers collect data in highly specific contexts -> tying data to standard vocabularies alone is insufficient and burdensome

Integration:

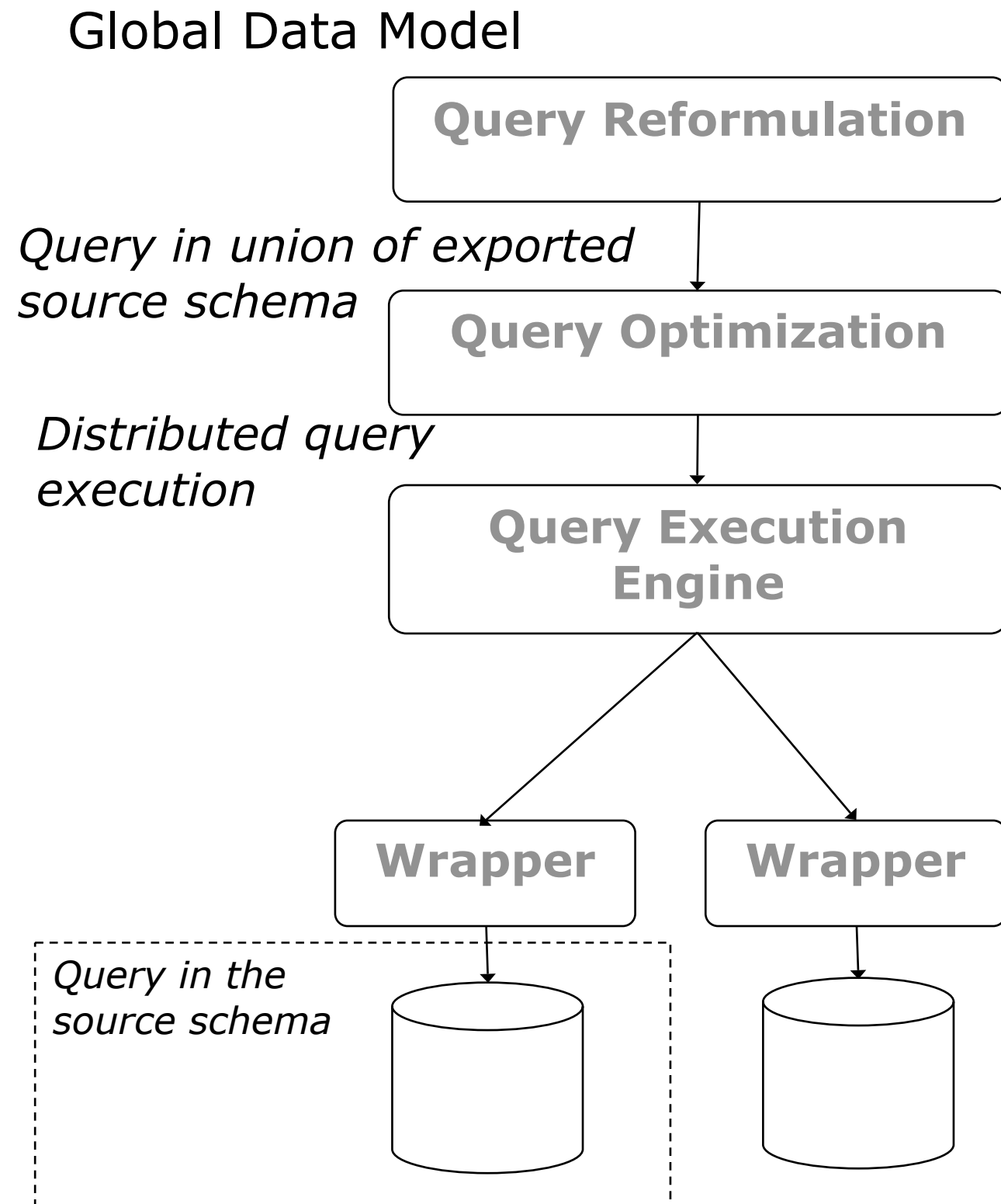
Generally used approaches

- Allow free text and lose interoperability
- Tightly encode data elements specific to purpose but lose expressivity/re-use and interoperability
- Post-hoc tying data elements to biomedical vocabularies
- Constraining choices to concepts in biomedical vocabularies
- Assemble raw data into warehouses

Integration via mediation

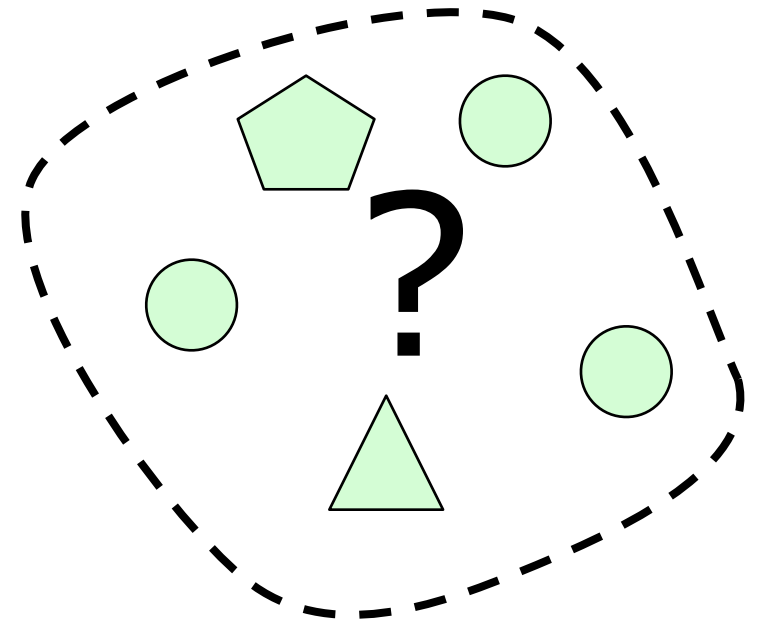
- Map between models
- Scoped to domain use
 - ◆ Multiple concurrent use
- Bottom up mediation
 - ◆ between standards and versions
 - ◆ between local versions
 - ◆ in absence of agreement

(Levy 2000)



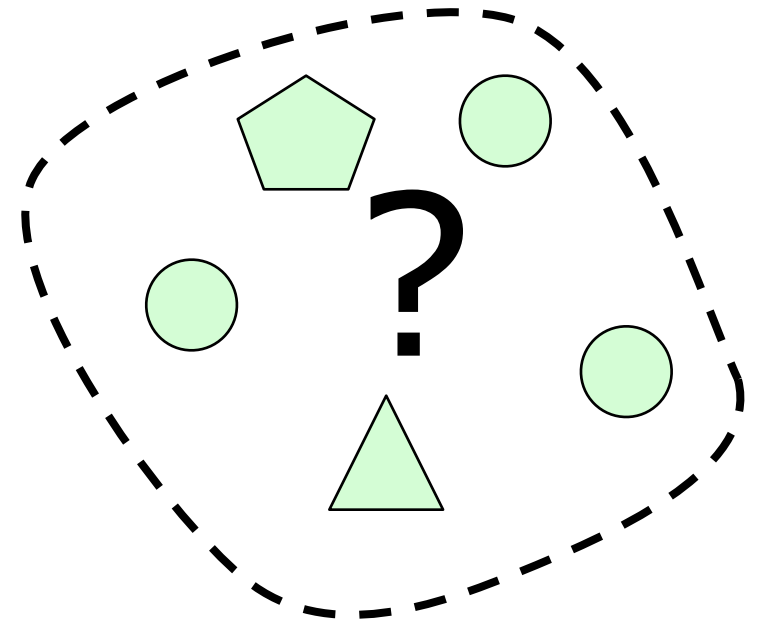
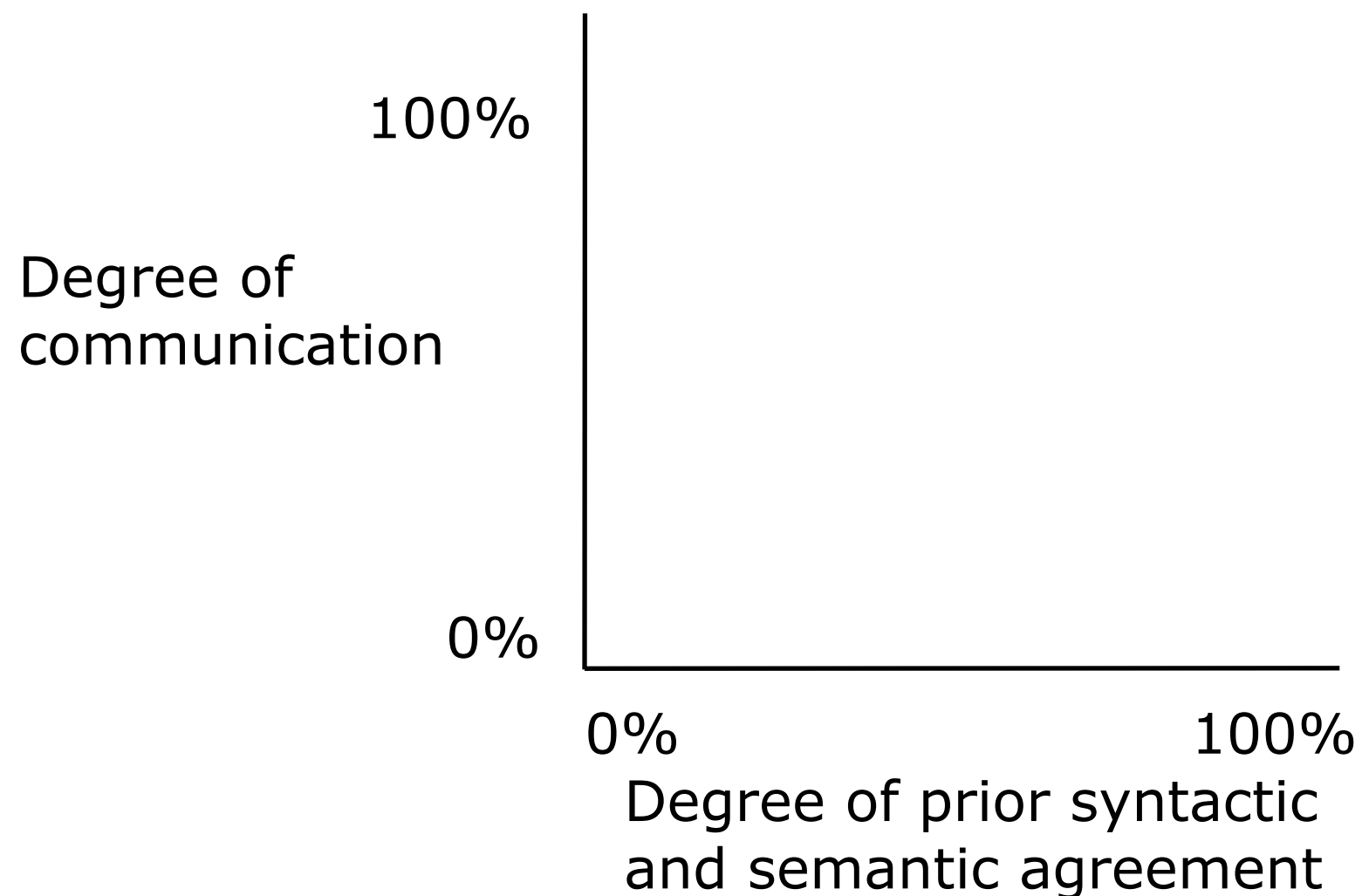
Integration:

Making data usable and useful



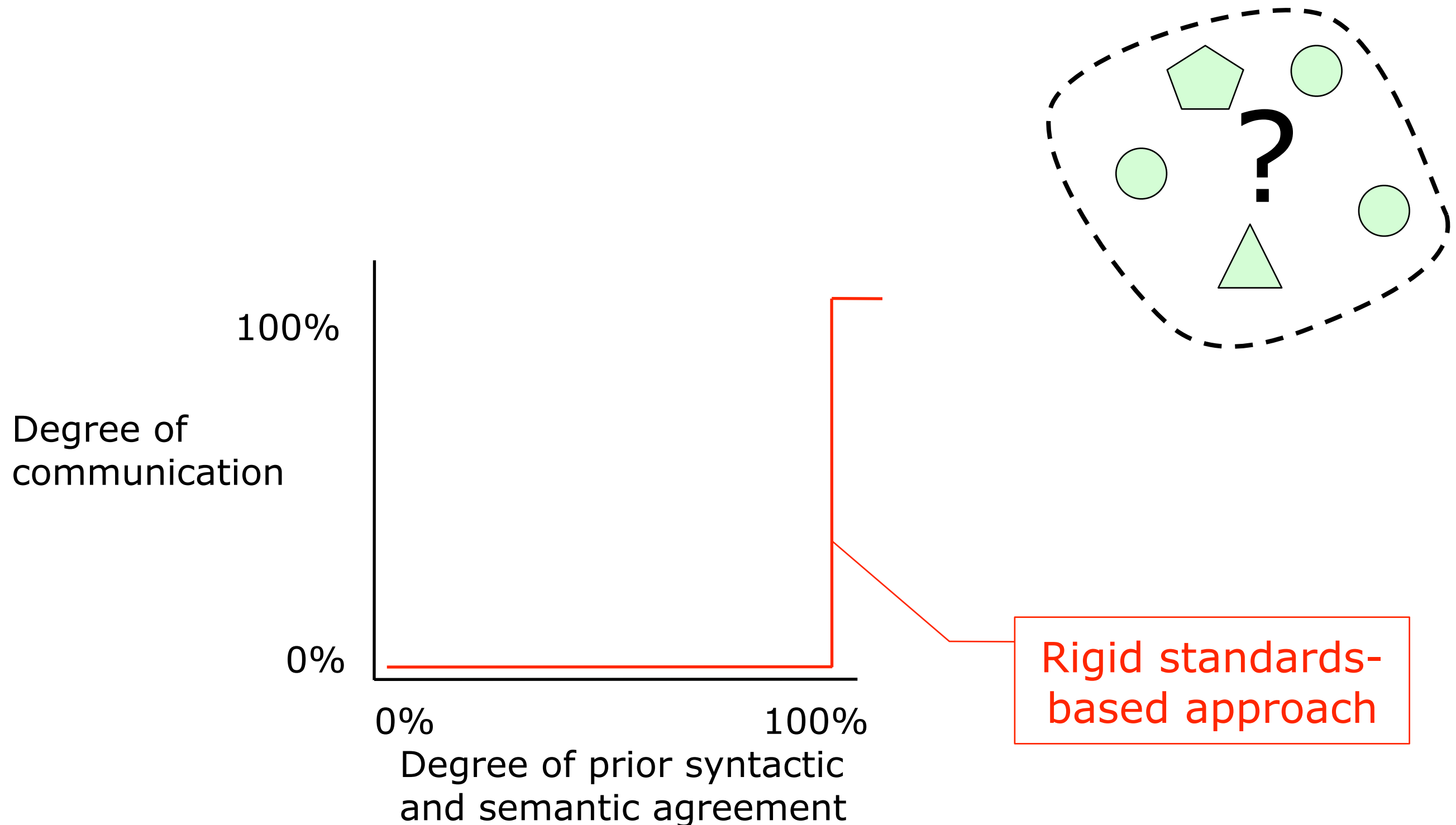
Integration:

Making data usable and useful



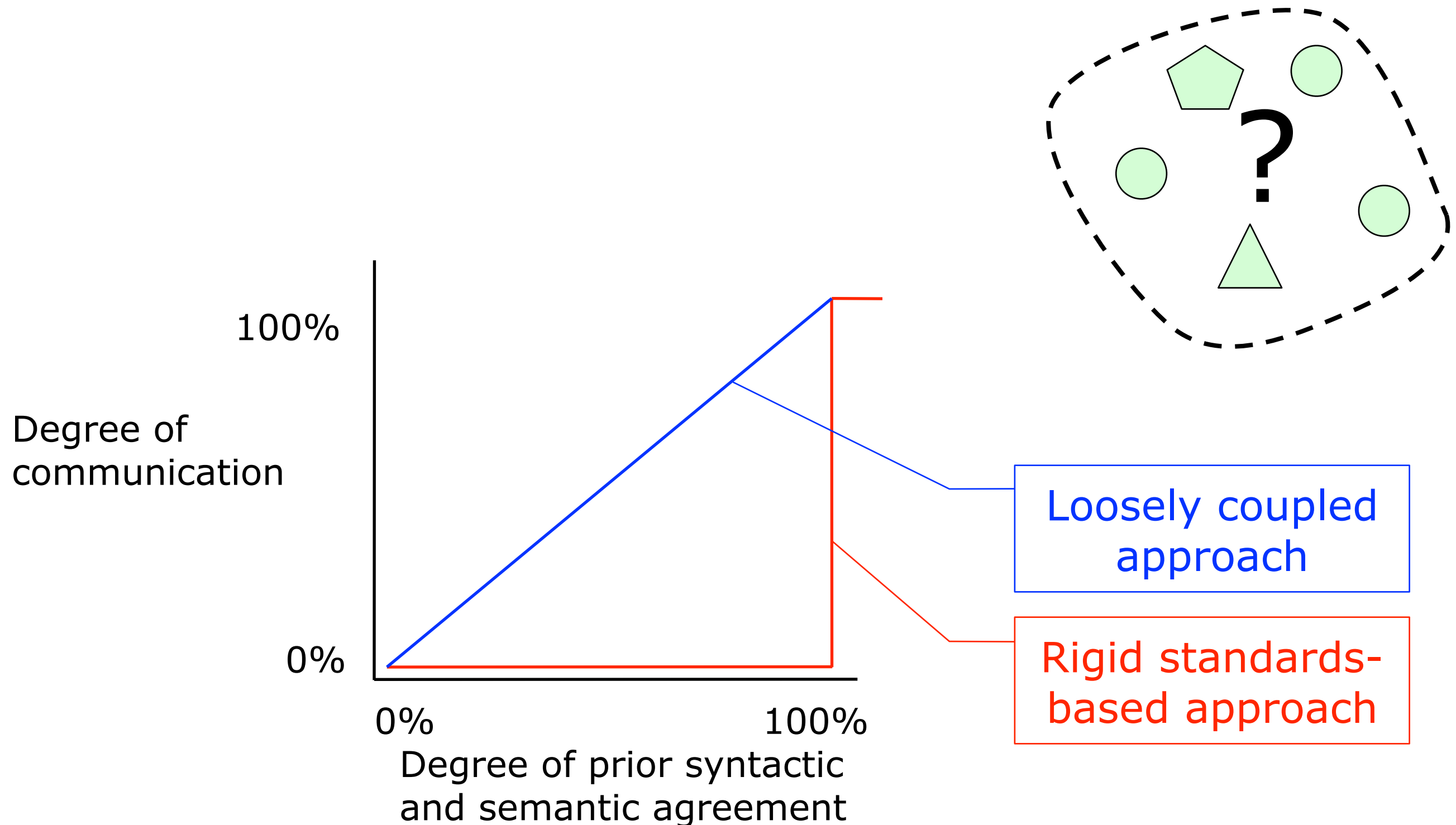
Integration:

Making data usable and useful



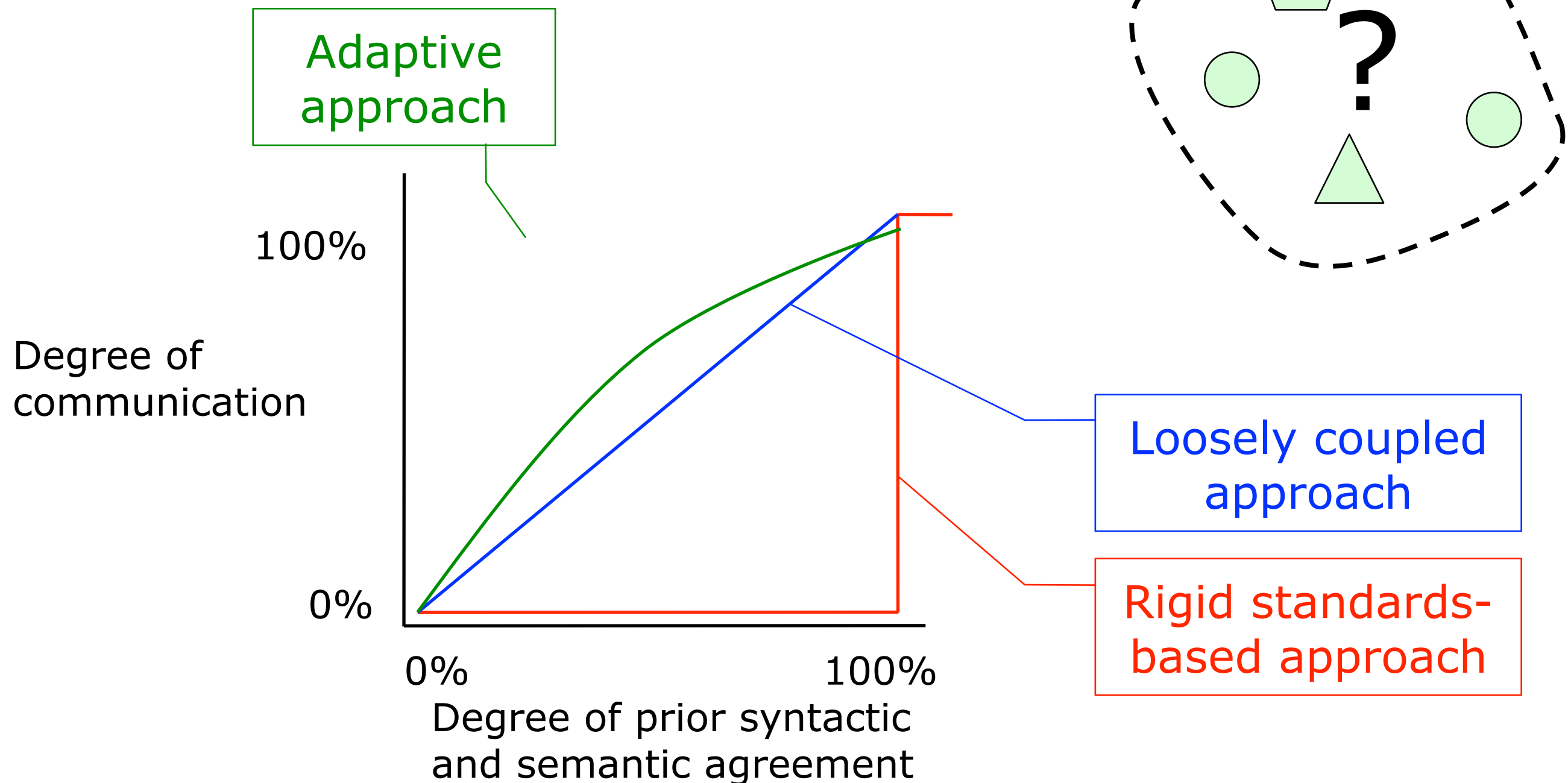
Integration:

Making data usable and useful



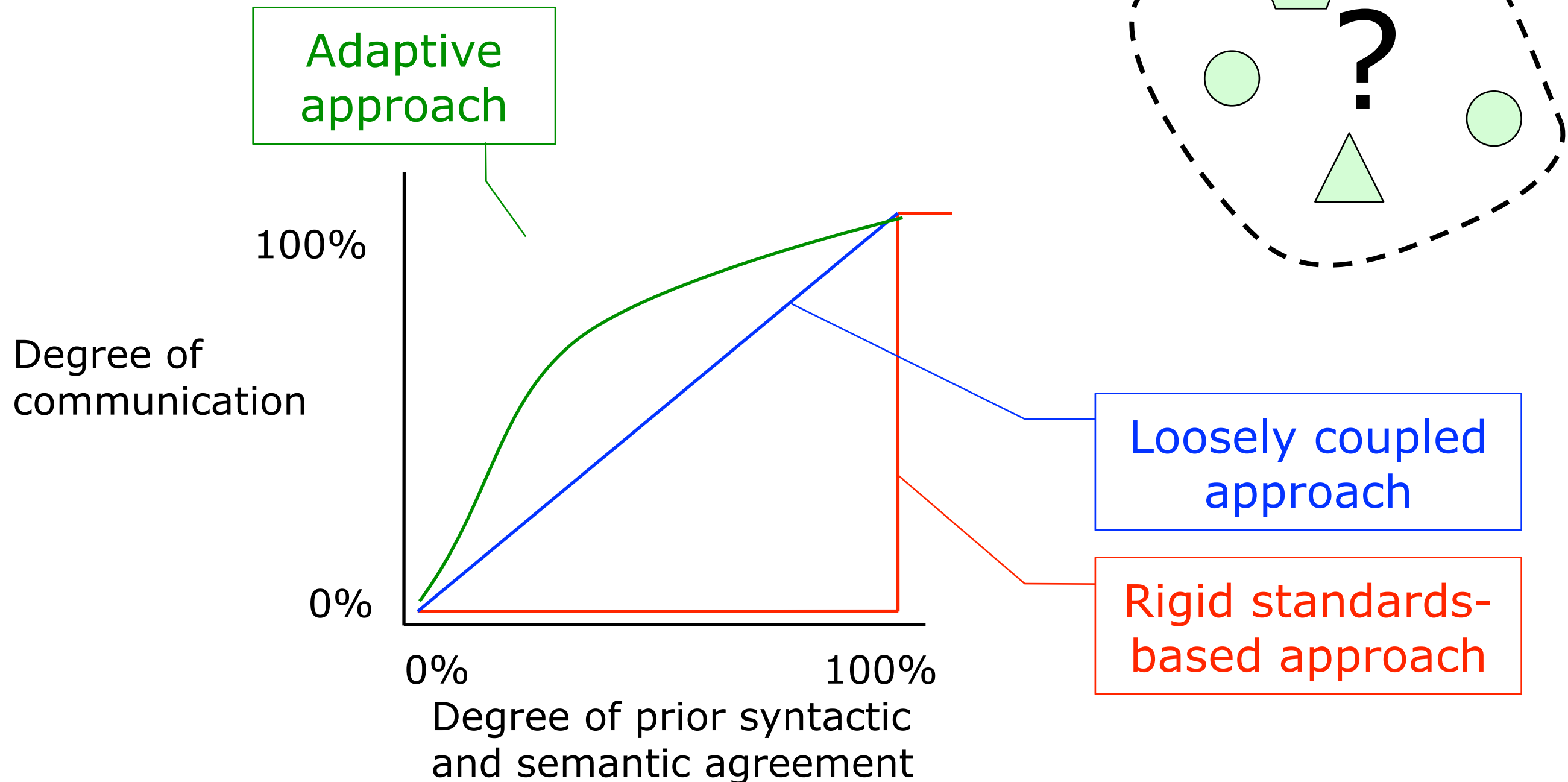
Integration:

Making data usable and useful



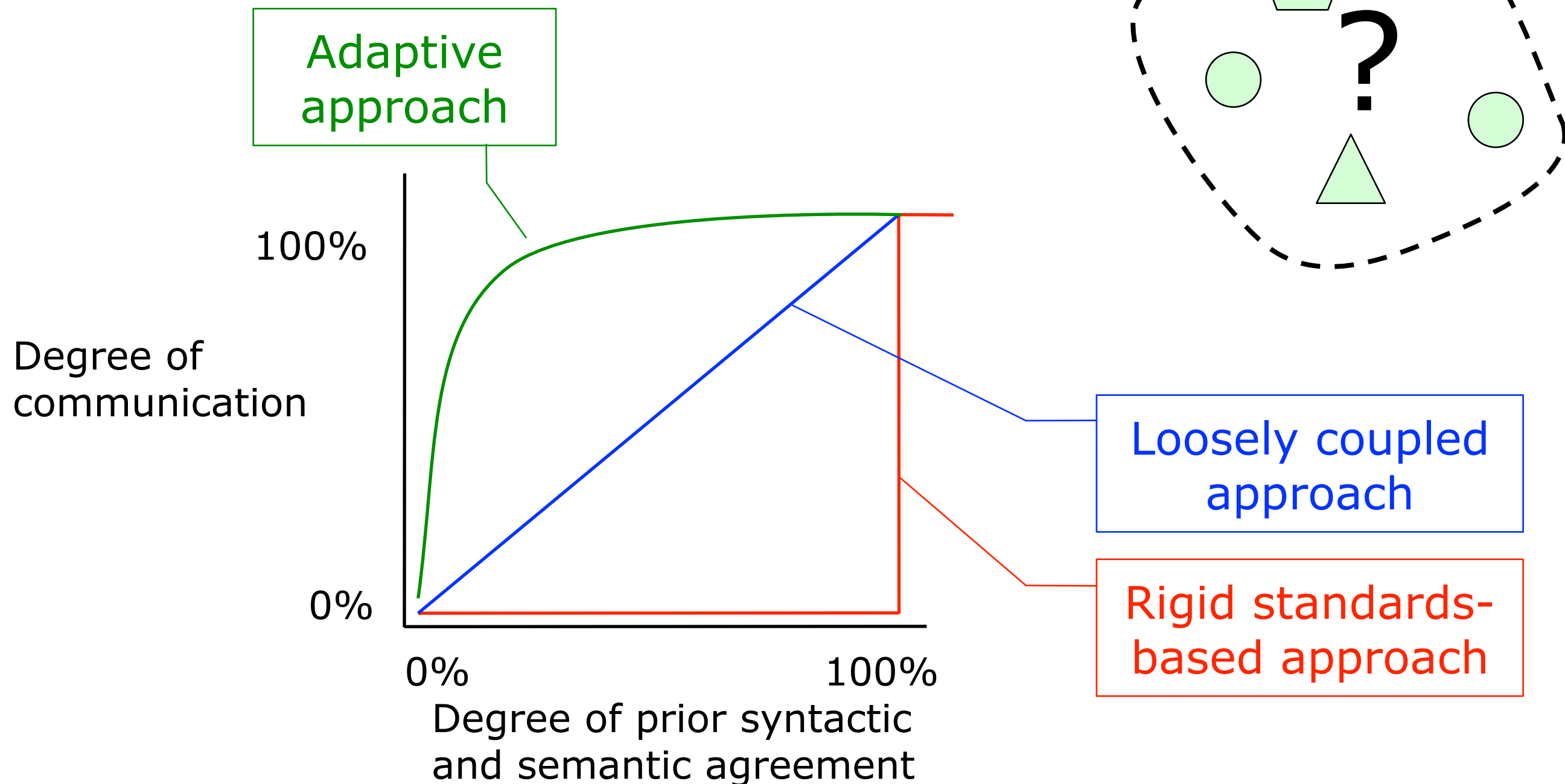
Integration:

Making data usable and useful



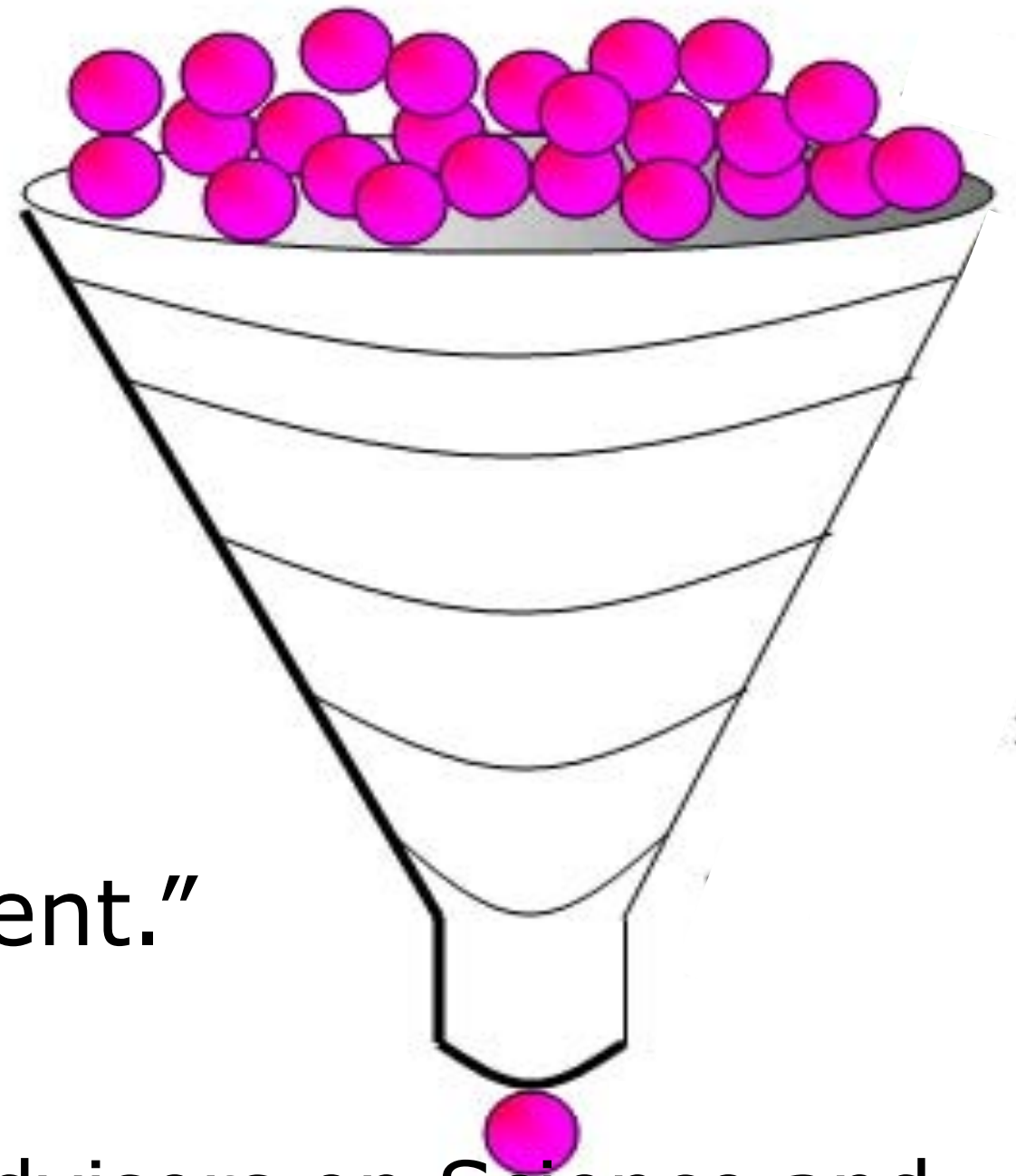
Integration:

Making data usable and useful



Analytics: Transform data into knowledge

“The overwhelming success of genetic and genomic research efforts has created an enormous backlog of data with the potential to improve the quality of patient care and cost effectiveness of treatment.”



— US Presidential Council of Advisors on Science and Technology, Personalized Medicine Themes, 2008

1. **Query** and retrieve microarray data from a caArray data service:

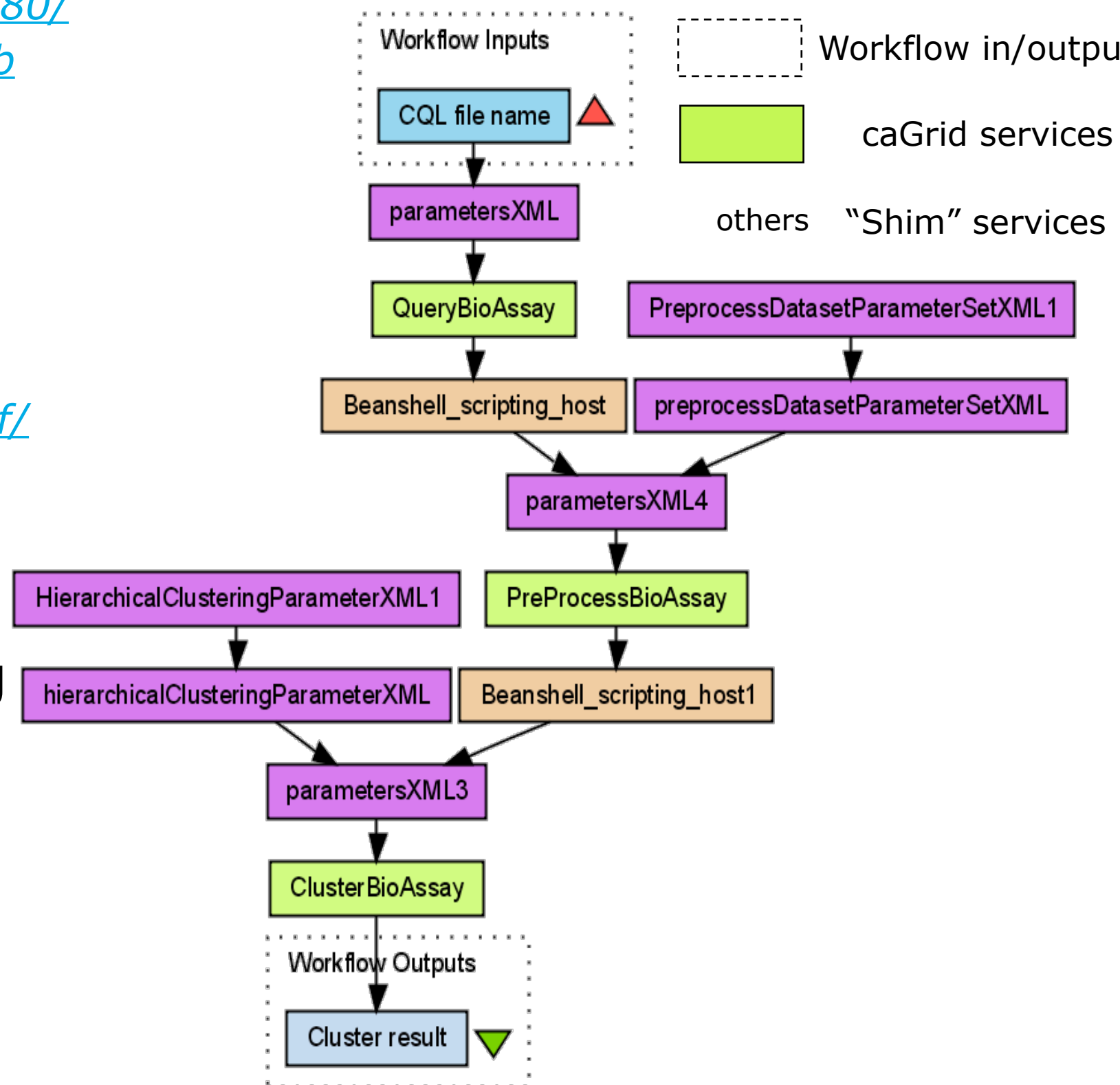
cagridnode.c2b2.columbia.edu:8080/wsrf/services/cagrid/CaArrayScrub

2. **Normalize** microarray data using GenePattern analytical service

node255.broad.mit.edu:6060/wsrf/services/cagrid/PreprocessDatasetMAGEService

3. **Hierarchical clustering** using geWorkbench analytical service:

Microarray clustering



The problem is understood

Health system is complex, adaptive system

There is no single point(s) of control. System behaviors are often unpredictable and uncontrollable, and no one is "in charge."

W Rouse, NAE Bridge

With diverse and evolving requirements and user communities

... I advocate ... a model of virtual integration rather than true vertical integration....

G. Halvorson, CEO Kaiser

Increased recognition that information systems and data understanding are limiting factor

... much of the promise associated with health IT requires high levels of adoption ... and high levels of use of interoperable systems (in which information can be exchanged across unrelated systems)

RAND COMPARE

Healthcare Transformation

Virtual organizations

+ hosted federation services

+ data integration

+ systems thinking and doing

+ large-scale computation

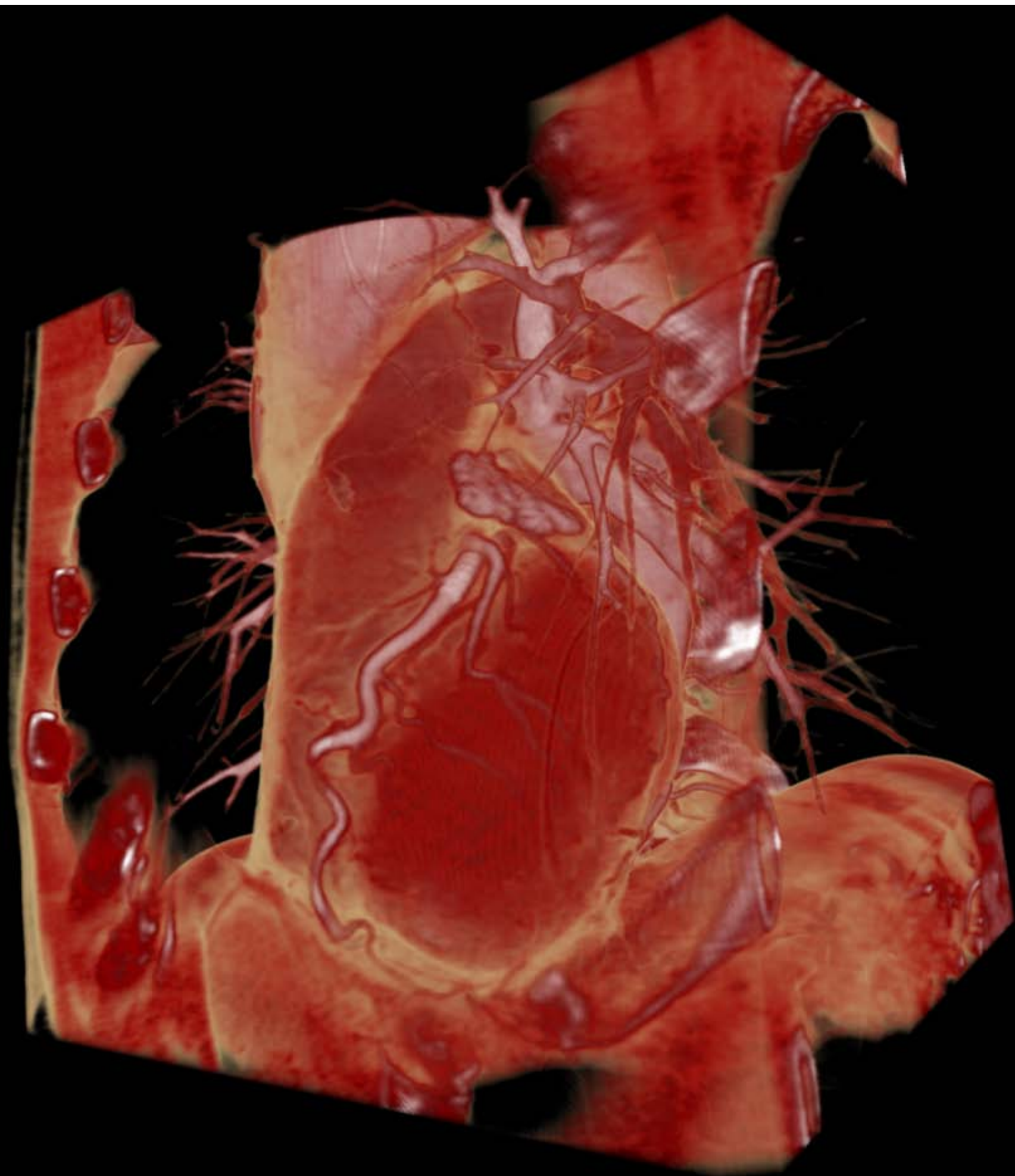
+ molecular medicine

= healthcare transformation

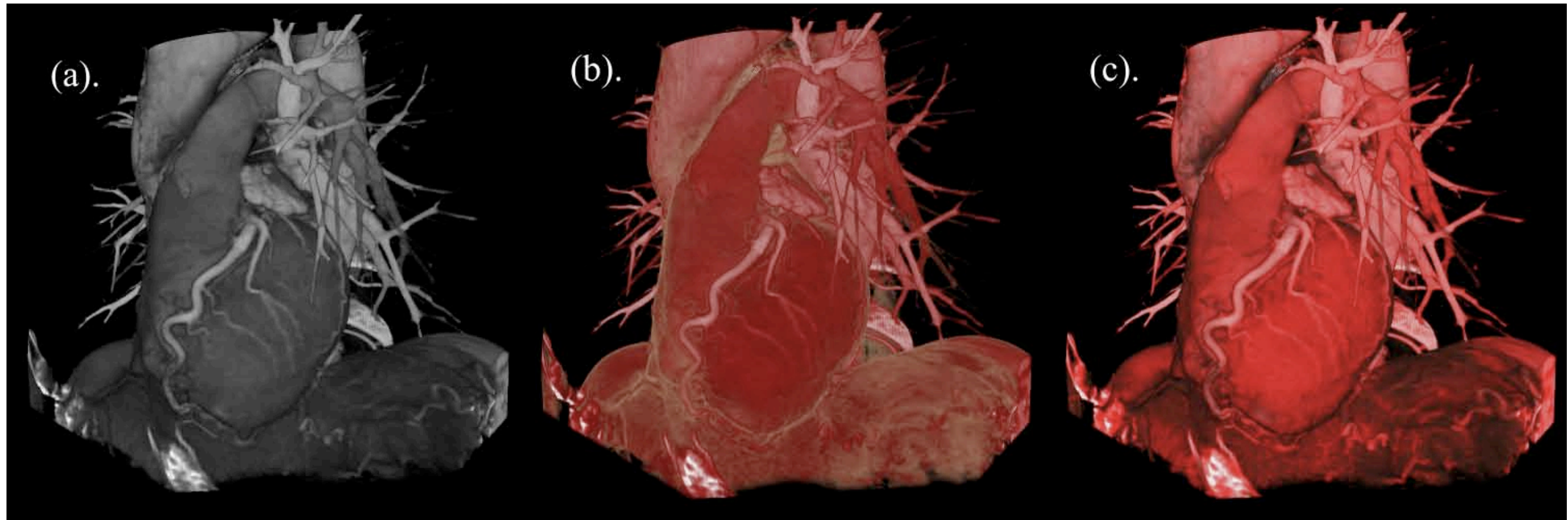
Synergies and Distinctions Between Computational Disciplines in Biomedical Research: Perspective From the Clinical and Translational Science Award Programs

Elmer V. Bernstam, MD, MSE, William R. Hersh, MD, Stephen B. Johnson, PhD, Christopher G. Chute, MD, DrPh, Hien Nguyen, MD, MAS, Ida Sim, MD, PhD, Meredith Nahm, MS, Mark G. Weiner, MD, Perry Miller, MD, PhD, Robert P. DiLaura, DBA, MBA, Marc Overcash, Harold P. Lehmann, MD, PhD, David Eichmann, PhD, Brian D. Athey, PhD, Richard H. Scheuermann, PhD, Nick Anderson, PhD, Justin Starren, MD, PhD, Paul A. Harris, PhD, Jack W. Smith, MD, PhD, Ed Barbour, MS, Jonathan C. Silverstein, MD, MS, David A. Krusch, MD, Rakesh Nagarajan, MD, PhD, and Michael J. Becich, MD, PhD, on behalf of the CTSA Biomedical Informatics Key Function Committee

“...the complementary but distinct roles of operational IT, research IT, computer science, and biomedical informatics”

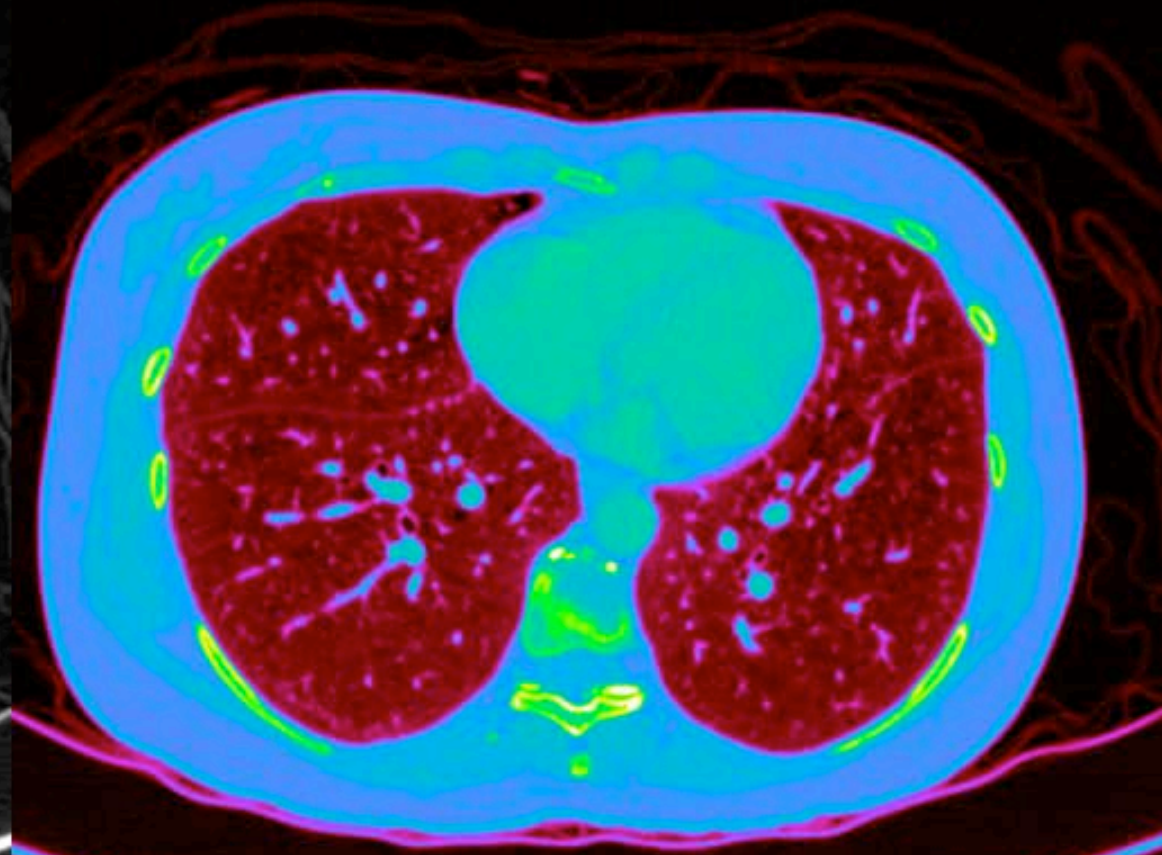
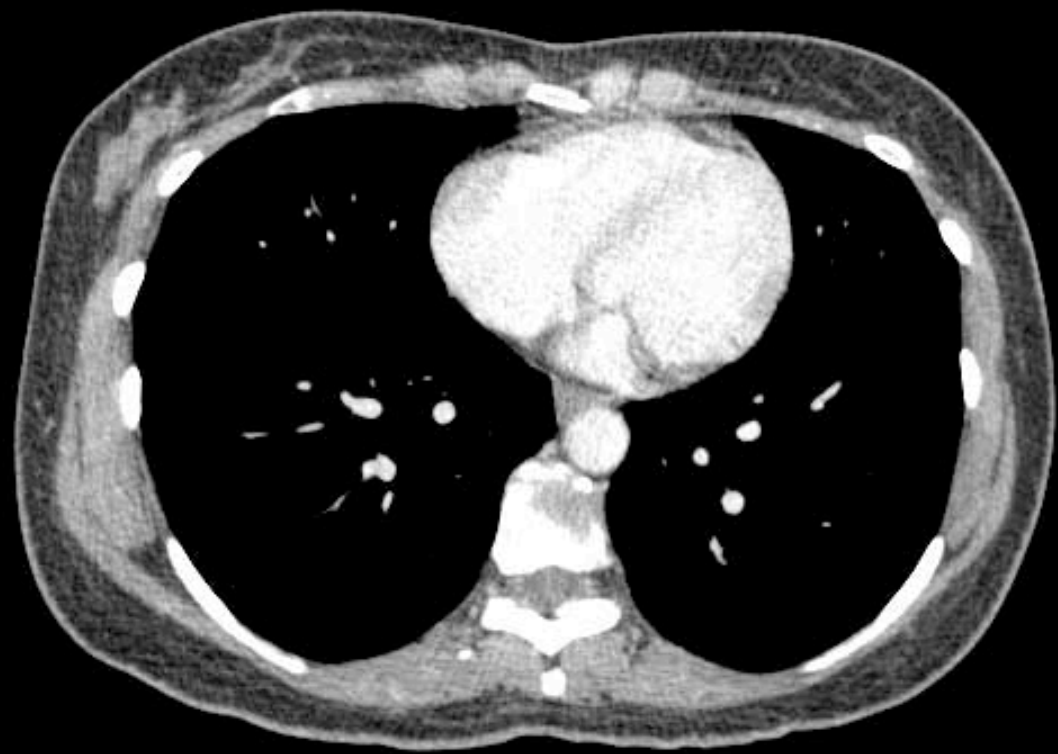
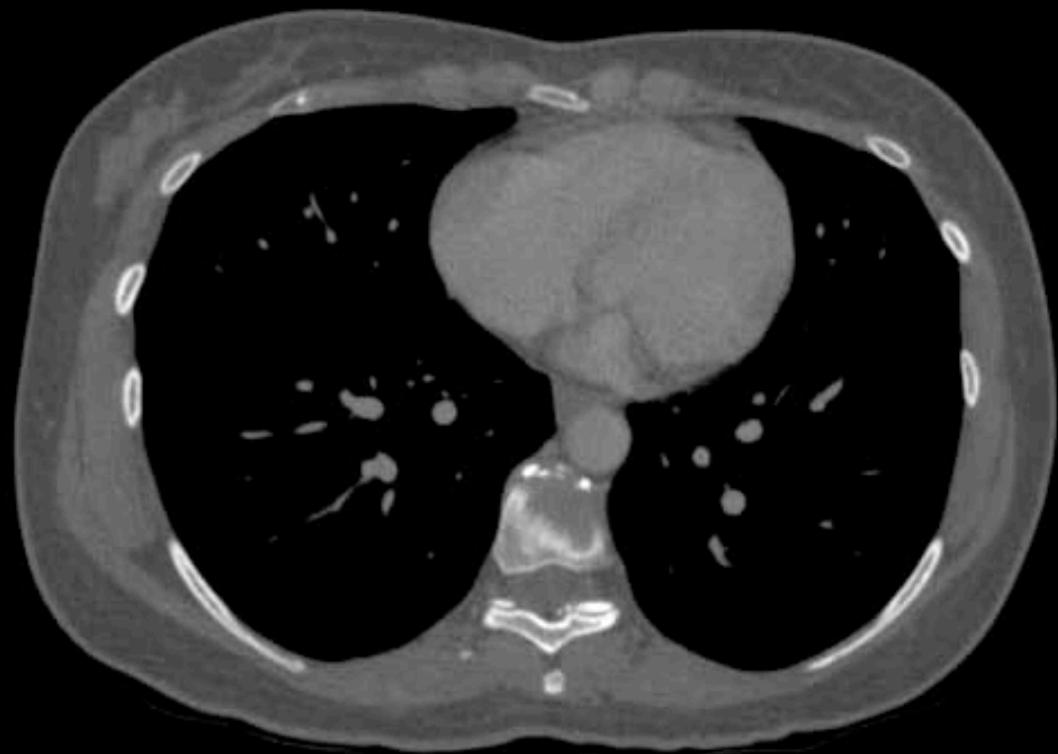


Automatic Perceptual Colormap Generation



Grayscale + Realistic Color = Perceptual

Silverstein JC, Parsad NM, Tsirline V. Automatic perceptual color map generation for realistic volume visualization. J Biomed Inform. 2008;41(6):927–935.



Volume Rendering with vl3

Refresh

Cancel

Refresh the display. If there are currently no valid VL3 requests, either in the queue or running, you will return to the Launch screen.

Cancelling will remove the request if vl3 has not yet started, or kill it on the TGviz nodes if it has. If it is running, it may take several seconds for it to be removed from the system, so you may need to press the "Refresh" button after canceling.

Command Output

vl3 server listening on host=tg-v049.uc.teragrid.org port=50111

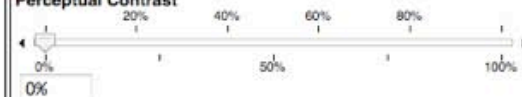


Motion Modes: ☐ Rotate ☐ Zoom ☐ Pan

Windowing: ☒ Bone

Rendering Mode: ☐ Gray ☒ Natural ☐ Spectral ☐ Thermal

Perceptual Contrast

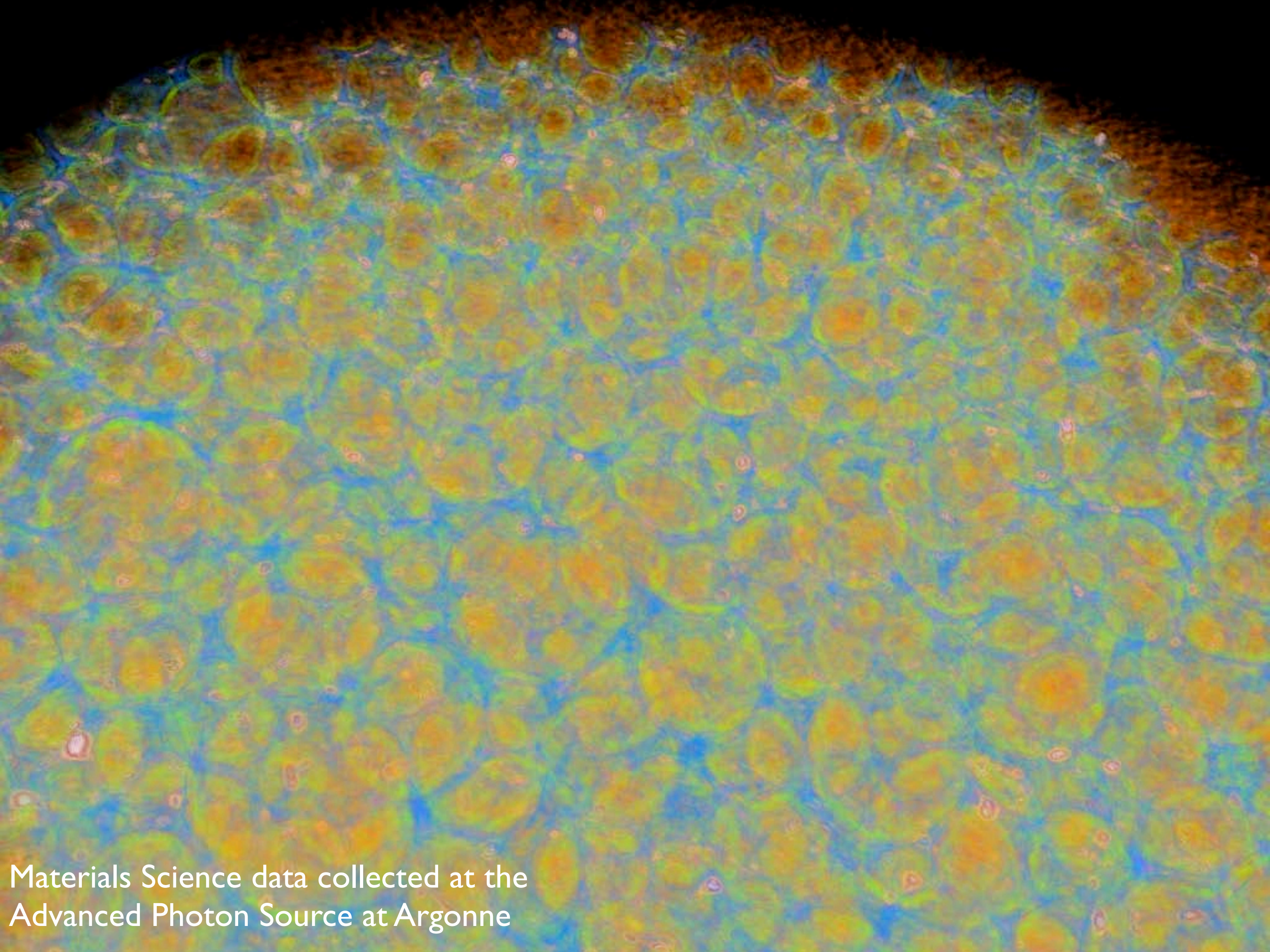


☐ Clipping:

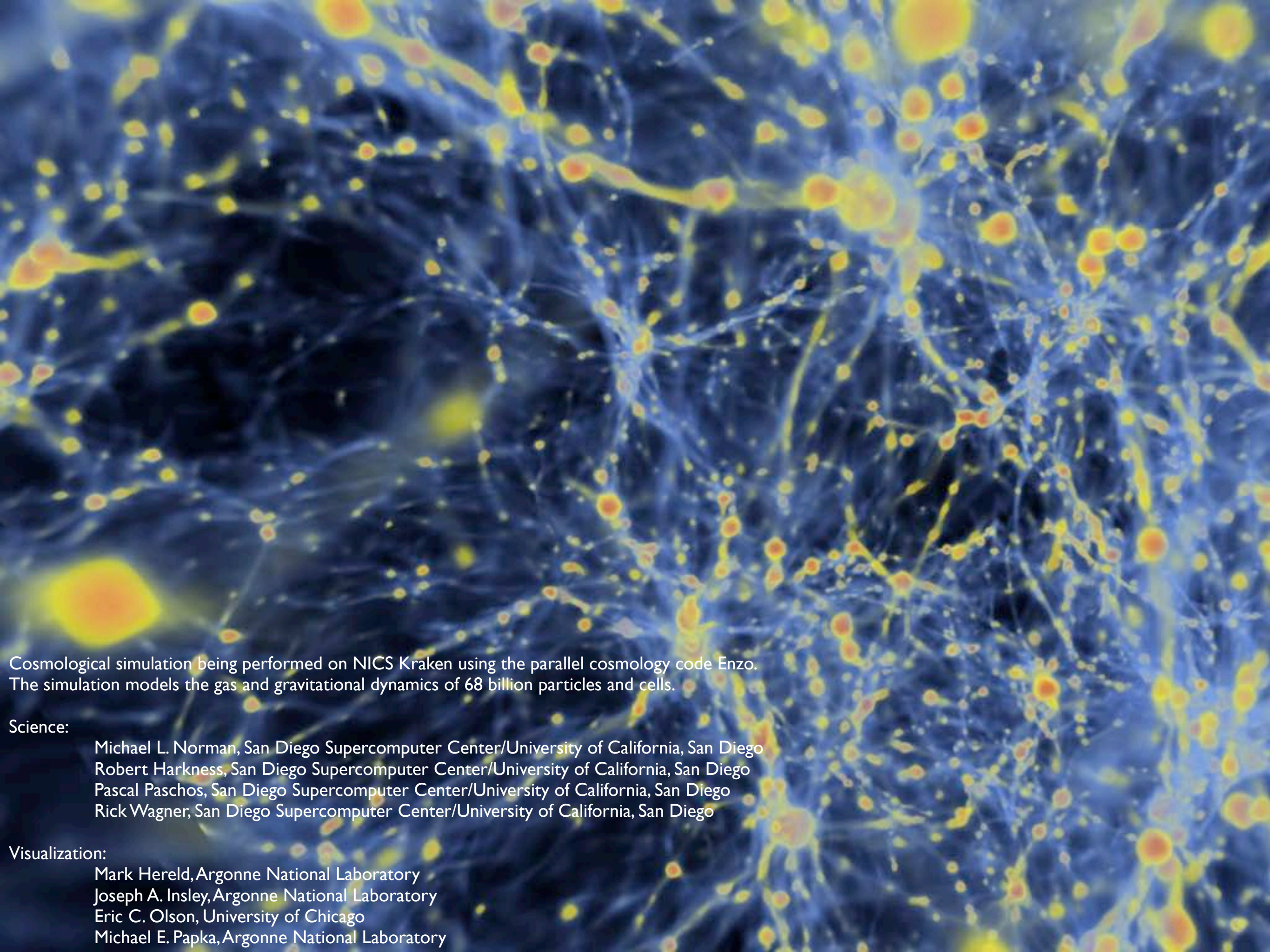
Sampling Freq:

Take a Snapshot





Materials Science data collected at the
Advanced Photon Source at Argonne



Cosmological simulation being performed on NICS Kraken using the parallel cosmology code Enzo.
The simulation models the gas and gravitational dynamics of 68 billion particles and cells.

Science:

Michael L. Norman, San Diego Supercomputer Center/University of California, San Diego
Robert Harkness, San Diego Supercomputer Center/University of California, San Diego
Pascal Paschos, San Diego Supercomputer Center/University of California, San Diego
Rick Wagner, San Diego Supercomputer Center/University of California, San Diego

Visualization:

Mark Hereld, Argonne National Laboratory
Joseph A. Insley, Argonne National Laboratory
Eric C. Olson, University of Chicago
Michael E. Papka, Argonne National Laboratory